THE SEVENTEENTH YEARBOOK

OF THE

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

PART II

THE MEASUREMENT OF EDUCATIONAL PRODUCTS

BY

E. J. ASHBAUGH, W. A. AVERILL, L. P. AYRES, F. W. BALLOU, EDNA BRYNER
B. R. BUCKINGHAM, S. A. COURTIS, M. E. HAGGERTY, C. H. JUDD
GEO. MELCHER, W. S. MONROE, E. A. NIFENECKER
E. L. THORNDIKE

Representing the
NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

Edited by GUY MONTROSE WHIPPLE, *Secretary*

This Part of the Yearbook will be discussed at the Atlantic City Meeting of the National Society, Monday, February 25th, 1918, 8:00 P. M. It will also be discussed at the Atlantic City meeting of the National Association of Directors of Educational Research

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS
1918

Published February, 1918

Composed and Printed by
Public School Publishing Company
Bloomington, Illinois.

OFFICERS OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

President

LOTUS D. COFFMAN
University of Minnesota, Minneapolis, Minnesota

Vice-President

J. A. C. CHANDLER
Superintendent of Schools, Richmond, Virginia

Secretary-Treasurer

GUY MONTROSE WHIPPLE
University of Illinois, Urbana, Illinois

Executive Committee

(THE YEAR INDICATES DATE OF EXPIRATION OF TERM)

HARRY B. WILSON (1918)
Superintendent of Schools, Topeka, Kansas

DWIGHT B. WALDO (1919)
State Normal School, Kalamazoo, Michigan

H. LESTER SMITH (1920)
Indiana University, Bloomington, Indiana

ERNEST HORN (1921)
University of Iowa, Iowa City, Iowa

Board of Trustees

S. CHESTER PARKER (1918)
University of Chicago, Chicago, Illinois
EDWARD C. ELLIOTT (1919)
University of Montana, Helena, Montana
GEORGE MELCHER (1920)
Bureau of Research and Efficiency, Kansas City, Missouri

TABLE OF CONTENTS

TABLE OF CONTENTS—Cont.

EDITOR'S PREFACE

In prefacing Part I of the Sixteenth Yearbook it was stated that the policy of the Society to give preference in its Yearbooks to contributions that would disseminate the reports of important committees in advance of the meetings at which they are to be discussed had been favorably received both by members of the Society and by large numbers of the educational public. It was also stated as "not unlikely that other committees and organizations of men professionally active in various aspects of educational endeavor will be glad to make similar use of the Society's avenues of publication in the future." This expectation has been realized in a most gratifying way in the Seventeenth Yearbook. I wish to congratulate members of the Society on this opportunity to serve themselves as well as the cause of education by cooperating with the National Association of Directors of Educational Research in the publication of the Yearbook these investigators have prepared.

G. M. WHIPPLE.

COMMITTEE'S PREFACE

The National Association of Directors of Educational Research present this Yearbook on Educational Measurement to the superintendents and teachers of American schools, hoping that it may prove of practical value to them in their work. Its purpose is to gather into one handy volume a rather complete statement of the various aspects of a new movement which seems destined to have a profound and permanent influence upon American education. Each chapter has been written by a different member of the Association, and as in any new field of work, complete agreement in either theory or practice is not to be expected, so in this volume the careful reader will detect many evidences of healthy variations in ideals, aims and methods. However, it is believed that these differences are not serious enough to mar the unity of plan and content and that the book as a whole represents the best judgment of the Association as to what information is of greatest practical worth. The Yearbook is issued in the hope that it may further the cause for which the Association stands—the promotion of educational research in American public schools.

The editorial committee, on behalf of the Association, hereby gratefully acknowledges its indebtedness to the National Society for the Study of Education, whose cooperation has made the publication of a yearbook possible.

EDITORIAL COMMITTEE

STUART A. COURTIS, *Chairman,*
LEONARD P. AYRES,
B. R. BUCKINGHAM,

CHAPTER I

HISTORY AND PRESENT STATUS OF EDUCATIONAL MEASUREMENTS

LEONARD P. AYRES
Division of Education, Russell Sage Foundation, New York City

Measurements in education are fifty years old if we count from the oldest beginnings of which we have record. They are twenty-five years old if we reckon from the time that Dr Rice, the pioneer and pathmaker among American scientific students of education, began his work in this field. They are ten years old if we begin our count with the earliest efforts of Professor Thorndike, who is the father of the present movement.

We are indebted to Professor Thorndike for having discovered what is apparently the earliest record of work in the field of educational measurements as we now use that term. As early as 1864 a school master in England, the Rev. George Fisher, of the Greenwich Hospital School, had seen the need and possibilities of standards, and with prophetic foresight anticipated present-day achievements. His practice was as follows: "A book called the 'Scale-Book' has been established, which contains the numbers assigned to each degree of proficiency in the various subjects of examination: for instance, if it be required to determine the numerical equivalent corresponding to any specimen of 'writing,' a comparison is made with the various standard specimens, which are arranged in this book in order of merit; the highest being represented by the number 1, and the lowest by 5, and the intermediate values by affixing to these numbers the fractions ¼, ½, or ¾. So long as these standard specimens are preserved in the institution, so long will instant numerical values for proficiency in 'writing' be maintained. And since fac-similes can be multiplied without limit, the same principle might be generally adopted.

"The numerical values for 'spelling' follow the same order and are made to depend upon the percentage of mistakes in writing

from dictation sentences from works selected for the purpose, examples of which are contained in the 'Scale-Book,' in order to preserve the same standard of difficulty.

"By a similar process values are assigned for proficiency in mathematics, navigation, Scripture, knowledge, grammar and composition, French, general history, drawing, and practical science, respectively. Questions in each of these subjects are contained in the 'Scale-Book,' to serve as types, not only of the difficulty, but of the nature of the questions, for the sake of future reference; observing that the same numerals are used in the same order as before, viz., number 1 denotes the highest, and number 5 the lowest amount of attainment.

"In respect to the numerical values of 'reading,' as regards accuracy, taste or judgment, it is obvious that no other standard of measurement can be applied, beyond the interpretation of the terms 'good,' 'bad,' 'indifferent,' etc., existing at the period of examination. And the same observation will apply to the estimation of numbers of 'characters' and 'natural abilities,' as determined by the united testimony of the respective masters.

"Having stated this much with regard to the plan pursued in this school, I may well add that the advantage derived from this numerical mode of valuation, as applied to educational subjects, is not confined to its being a concise method of registration, combined with a useful approximation to a *fixed standard* of estimation, applicable to each boy; but it affords also a means of determining the *sum total,* and therefrom the means or average condition or value of any given number of results.'"[1]

Mr. Fisher's efforts seem to have produced no lasting results. Progress in the scientific study of education was not possible until people could be brought to realize that human behavior was susceptible of quantitative study, and until they had statistical methods with which to carry on their investigations. Both of these were contributed in large measure by Sir Francis Galton. As early as 1875 he published scientific studies of the traits of twins, of number-

[1]Reported by E. B. Chadwick in the *Museum, a Quarterly Magazine of Education, Literature and Science,* Vol. III, 1864. See also *Journal of Educational Psychology,* Vol. 4, page 551.

forms, of color-blindness, and of the efficacy of prayer. Out of his work came much of experimental and educational psychology, and indirectly, educational measurements. It was he who developed the statistical methods necessary for the quantitative study of material which seemed at the outset entirely qualitative and not at all numerical in nature.

In America the real inventor of the comparative test was Dr. J. M. Rice. Dr. Rice studied in Germany and came under the influence of the German psychologists at Jena and Leipsic. Returning to this country, he became interested in education and one day in 1894 the new idea was born. Of this invention Dr. Rice says: "In truth, however, I came to recognize that this (the claims of school men following different courses of study) was all talk,—that no one really knew the facts, because there were no standards to serve as guides. Then one day, the idea flashed through my mind that the way to settle the question was to try it out. For a beginning I decided to take spelling, and on that very day I made up a list of 50 words with the view of giving them as a test to the pupils of the schools as I went on my tour from town to town. I have no record of the date of the inspiration, but I think it was some time in October, 1894."

Dr. Rice's work, however, did not meet with the approval of the educators of the day. One of his earlier reports in this field indicated that children who had spent thirty minutes a day for eight years in the study of spelling did not at the end of that time spell any better than the children in another school system who had spent only fifteen minutes a day for eight years in the same study. The presentation of these results brought upon the investigator almost unlimited attack. The educators who discussed his findings and those who reviewed them in the educational press united in denouncing as foolish, reprehensible, and from every point of view indefensible, the effort to discover anything about the value of the teaching of spelling by finding out whether or not the children could spell. They claimed that the object of such work was not to teach children to spell, but to develop their minds! It was the issue between the investigator and the formalist in education, and the con-

flict that is still under way is the conflict that was then for the first time clearly defined.

Little by little the more thoughtful men in the field of education appreciated the suggestive value of Dr. Rice's work and some few of them, notably Professor Hanus of Harvard, dared to come to his support. Slowly the tide turned in his favor, until by common consent the general validity of his conclusions was tentatively accepted. His methods, however, were not generally adopted, and for more than ten years but little progress was made beyond the work of the pioneer himself.

If Dr. Rice is to be called the inventor of educational measurement, Professor E. L. Thorndike should be called the father of the movement. In 1895, Professor Thorndike was a student at Columbia, struggling with statistical methods in a course on measurements under Boas, and "finding it new and very hard for me to learn."[2] His interests were in the field of psychology and the work of Rice made a deep and lasting impression. Gradually his experimental work came more and more into the educational field. He began to preach the need of measurement and to experiment with tests and scales. The Stone Arithmetic Tests were published in 1908. The Thorndike Scale for the measurement of merit in handwriting was presented before Section L of the American Association for the Advancement of Science at its Boston meeting, in December, 1909, and was published in the *Teachers College Record* the following March. The construction of this scale, based on the equal difference theorem formulated by Cattell, marks the real beginning of the scientific measurement of educational products.

During the past ten years the growth of the scientific movement in education has been continuous and rapid. It has been closely related to the survey movement which had its real inception in 1907 in a great social study of the city of Pittsburgh, which was termed "A Survey." Three years later two college professors, Hanus of Harvard and Moore of Yale, conducted studies of the school systems of Montclair and East Orange in New Jersey. These studies differed from earlier investigations of school systems in that their purpose was to tell the public about their public schools, and each

[2]Quotation from a personal letter.

investigator, borrowing the term from the contemporary social movement, used the word "survey" to designate a section of his report. In the years that have followed, scores of surveys of city, state and county school systems have been conducted, and in ever increasing degree they have utilized, as perhaps the most important of their methods, the scales and tests used in the measurement of educational processes and products.

The two movements were represented in the New York school inquiry of 1911-12. For the first time in a formal educational investigation tests were used as an aid in evaluating the results of public-school work. These were the Courtis arithmetic tests, which had by that time attracted a good deal of attention. Their successful use in the New York survey not only settled all doubts as to the availability of the tests themselves for the measurement of educational attainment, but also firmly established the principle that in conducting school surveys scientific tests must be utilized where they are available.

One of the recommendations of the survey committee in New York City was that a Bureau of Research be established to conduct a continuous survey, from within the school system itself and for its benefit. This recommendation was immediately adopted and the Bureau organized in September, 1913. Previous to this time, there had been in various cities committees and other organizations, which had made studies of various phases of administrative and instructional work, but to New York City probably belongs the credit of first establishing a formal organization having for its purpose the continuous critical study of its own activities by scientific methods.

By this time, Boston, Detroit, and many other cities, were experimenting with measurement and obtaining results of value. Other bureaus were soon established. The Division of Education of the Russell Sage Foundation had turned its attention to work of this type as early as 1907, the Boston Bureau was organized in 1913, and similar organizations in Detroit, Kansas City, and Oakland soon followed. During these same years faith and interest in measurements had been greatly stimulated by the development of the Binet-Simon tests and by the wide-spread attention given the study of retardation and elimination in school systems. A demand for

men trained for the work was created. Superintendents and teachers also were clamoring for technical knowledge of methods and for explanations of the results obtained. For ten years graduates of Teachers College, Columbia University, and of the School of Education of Chicago University had had impressed upon them that measurement and scientific experimentation were highly desirable in education and they had been at least partly prepared for such work. Next, this training was expanded into formal courses in educational measurement. Soon courses in measurement appeared in all the great universities and the movement began to gain full momentum.

The meetings of the Department of Superintendence of the National Education Association afford an excellent index of the progress of the movement. Dr. Rice's report in 1897 was received with derision. The Philadelphia meeting in 1912, after a heated discussion, voted against measurement by a small majority, but two years later a committee on Tests and Standards made a favorable report which was adopted by a considerable majority.

Today tests and scales are used throughout this country and around the world. In England, Germany, and France, before the war, beginnings had been made. Scales for the measurement of Chinese writing and composition are now in process of construction. In Australia and New Zealand, in India and Hawaii, and throughout the length and breadth of the United States and Canada, tests and scales are in daily service, proving valuable tools in the hands of those who know how to use them.

The scientific method is at base analytic scrutiny, exact measuring, careful recording, and judgment on the basis of observed fact. Science in education is not a body of information, but a method, and its object is to find out and to learn how. By its aid, education is becoming a profession. Courses of study are being adapted to the needs of children; teaching effort and supervisory control are becoming more efficient. The center of interest in education has become the child, rather than the teacher, and efforts to improve the quality of instruction begin by finding out what the children can do, rather than by discussing the methods by which the teacher proceeds.

Educational measurement has been accepted by the American public. This Yearbook is in itself a proof of that. That the methods of today are still crude and imperfect must be admitted by even the most enthusiastic supporters of the movement. They deal most effectively with only the simple mechanical skills, and even here they are still far from perfect. Nevertheless, they are extending each year their range of availability and their field of application.

The importance of the movement lies not only in its past and present achievements, but in the hope of the future. Knowledge is replacing opinion, and evidence is supplanting guess-work in education as in every other field of human activity. This is the supreme fact to which this Yearbook bears witness. The future depends upon the skill, the wisdom, and the sagacity of the school men and women of America. It is well that they should set about the task of enlarging, perfecting, and carrying forward the scientific movement in education, for the great war has marked the end of the age of haphazard, and the developments of coming years will show that this is true in education as in every other organized field of human endeavor.

CHAPTER II

THE NATURE, PURPOSES, AND GENERAL METHODS OF MEASUREMENTS OF EDUCATIONAL PRODUCTS

EDWARD L. THORNDIKE
Professor of Educational Psychology, Teachers College, Columbia University

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality. Education is concerned with changes in human beings; a change is a difference between two conditions; each of these conditions is known to us only by the products produced by it—things made, words spoken, acts performed, and the like. To measure any of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. To measure a product well means so to define its amount that competent persons will know how large it is, with some precision, and that this knowledge may be conveniently recorded and used. This is the general *Credo* of those who, in the last decade, have been busy trying to extend and improve measurements of educational products.[1]

We have faith that whatever people now measure crudely by mere descriptive words, helped out by the comparative and superlative forms, can be measured more precisely and conveniently if ingenuity and labor are set at the task. We have faith also that the objective products produced, rather than the inner condition of the person whence they spring, are the proper point of attack for the measurer, at least in our day and generation.

This is obviously the same general creed as that of the physicist or chemist or physiologist engaged in quantitative thinking—the same, indeed, as that of modern science in general. And, in general,

[1]The conditions to be thoroughly known must be known as quantities a, b, c, d, etc., of qualities, or powers, or skills, or knowledges A, B, Y, J, etc.—that is, as an equation $aA+bB+cY$, etc.

the nature of educational measurements is the same as that of all scientific measurements.

In detail, however, there are notable differences. An educational product, such as a composition written, a solution of a problem in arithmetic, an answer to a question about history, a drawing of a house or the performance of an errand, is commonly a complex of many sorts of things. The task of measuring it seems more like measuring a house or an elephant than it is like measuring a length or a volume or a weight. A complete measurement of, say, a composition might include an exact definition of its spelling, its usage of words, its usage of word forms, its wit, its good sense and so on and on; and each of these might again be subdivided into a score or more of component elements.

What we do, of course, is to make not such a complete measurement of the total fact, but to measure the amount of some feature, *e.g.*, the general merit of the composition or the richness of its vocabulary, just as physical science does not measure the elephant, but his height, or his weight, or his health, or his strength of pull. Every measurement represents a highly partial and abstract treatment of the product. This is not understood by some of our critics who object to tests and scales because of their limited point of view. The critic's real point should be that an educational product commonly invites hundreds of measurements, as we all well know. It should be noted also that single measurements are still in a sense complex, being comparable to volume, wattage or the opsonic index, rather than to length, weight or temperature.

In the second place, the zeros of the scales for the educational measures and the equivalence of their units are only imperfectly known. As a consequence, we can add, subtract, multiply and divide educational quantities with much less surety and precision than is desirable. Indeed, in any given case, the sense in which one educational product is twice as good or as desirable as another, or in which one task is twice as hard as another, or in which one improvement is twice as great as another, is likely to be a rather intricate and subtle matter, involving presuppositions which must be kept in mind in any inferences from the comparison.

In some cases so little is known of units of amount that we do not even try to equate distances along the scale, but simply express relative size in terms of arbitrarily chosen units and reference points.[2] This is the case, for example, with the most commonly used measurement in psychology and education, that due to applying the Binet-Simon tests.

Nobody need be disturbed at these unfavorable contrasts between measurements of educational products and measurements of mass, density, velocity, temperature, quantity of electricity, and the like. The zero of temperature was located only a few years ago, and the equality of the units of the temperature-scale rests upon rather intricate and subtle presuppositions. At least, I venture to assert that not one in four of, say, the judges of the supreme court, bishops of our churches, and governors of our states could tell clearly and adequately what these presuppositions are. Our measurements of educational products would not at present be entirely safe grounds on which to extol or condemn a system of teaching reading or arithmetic, but many of them are far superior to the measurements whereby our courts of law decide that one trademark is an infringement on another.

There are two somewhat distinct groups of educational measurements: one, well illustrated by the Courtis tests, asks primarily how well a pupil performs a certain uniform task; the other, well illustrated by the Hillegas or Trabue tests, asks primarily how hard a task a pupil can perform with substantial perfection, or with some other specified degree of success. The former are allied to the so-called method of average error of the psychologists; the latter, to what used to be called the method of "right and wrong cases." Each of these groups of methods has its advantages, and each deserves extension and refinement, though the latter seems to represent the type which will prevail if education follows the course of development of the physical sciences.

I have so far omitted specific reference to measurements by relative position—the so-called 'order of merit method.' This method, available even where no differences in the amount of the

[2] That is, as a, $a+x$, $a+x+y$, $a+x+y+z$, etc., claiming only that x, y, and z are all positive quantities.

thing measured are defined, and useful to organize the reports of untrained observers, is doing excellent service as a first stage in quantitative knowledge. For every reason, however, the grading of a set of educational products by relative position should soon give way to their rating by some even rough scale, such as meteorology uses for the cloudiness of days, and such as astronomers use for the magnitude of stars. A very, very simple form of scale, such as almost anybody can use to measure almost any product that he knows anything about, is that devised by Walter Dill Scott for use in rating the achievements or promise of employees.[3]

The purpose of measurements of educational products is in general to provide somebody with the knowledge that he needs of the amount of some thing, difference or relation. The "somebody" may be a scientific worker, a superintendent of schools, a teacher, a parent or a pupil. He may need a very precise or only an approximate measure, according to the magnitude of the difference which he has to determine. He may need it for guidance in many different sorts of decisions and actions.

Some of the most notable uses concern the values of studies in terms of the changes produced by them, the effects of different methods of teaching, and the effects of various features of a school system, such as the salary scale, the length of the school day and year, the system of examining and promoting pupils, or the size of class. There are many problems under each of these heads, and each of these problems is multifarious according to the nature, age, home life and the like of the pupils, and according to the general constitution of the educational enterprise, some small feature of which is being studied.

Another important group of uses concerns inventories of the achievements of certain total educational enterprises such as our educational surveys must become if they are to carry authority with scientific men. The total educational enterprise may be the work

[3]The Rating Scale was originally devised and is now issued as one of the copyrighted forms of the Bureau of Salesmanship Research, Carnegie Institute of Technology, Pittsburgh, Pa. A modification of it is in use in certain sections of the military organization of the United States for the selection of men for promotion. A description of this military modification has appeared in *Collier's Weekly*—G. M. W.

of a teacher, of a school, of an orphanage, of a prison, of a system of schools, or the like.

Another important group of uses centers around the problem of giving the individual pupil the information about his own achievement and improvement which he needs as a motive and a guide. It is interesting to note that the first of the newer educational scales, which was expected to be used chiefly by scientific investigators of the teaching of handwriting, now hangs on the wall of thousands of classrooms as a means for pupils to measure themselves. There are many other purposes, and important ones, such as the detection and removal of gross prejudices on the part of teachers in their own evaluations of certain educational aims and products. These,however, cannot be described here.

The superintendents, supervisors, principals and teachers directly in charge of educational affairs have been so appreciative of educational measurements and so sincere in their desire to have tests and scales devised which they can themselves apply, that the tendency at present is very strong to provide means of measurement which are concerned somewhat closely with school achievements, and which can be used by teachers and others with little technical training. There is also a tendency, because of this need for a large number of measurements in the case of educational problems, to try to devise tests which can be scored by persons utterly devoid of judgment concerning the products in question.

It would ill become the present writer to protest against these two tendencies; and they are intrinsically healthy. There is, however, a real danger in sacrificing soundness of principle and precision of result to the demand that we measure matters of importance and measure them without requiring elaborate technique or much time of the measurer. The danger is that the attention of investigators will be distracted from the problems of pure measurement for measurement's sake, which are a chief source of progress in measuring anything. Perhaps not even one person in a million need feel this passion, but for that one to cherish it and serve it is far more important than for him to devise a test which thousands of teachers will employ. Opposition, neglect, and misunderstanding will be much less disastrous to the work of quantitative science

in education than a vast output of mediocre tests for measuring this, that and the other school product, of which a large percent are fundamentally unsound.

We have seen that educational measurements vary from an assignment of a certain amount of some clearly defined thing, the zero, or "just not any," of which is fairly accurately known, to a mere assignment of a certain position in a series of products themselves only similarly defined. They vary also from measurements in the most unimpeachable of units, such as time, to measurements where the unit is "that difference in quality which 75 percent of a certain sort of observers succeed in observing" or is even more crudely and hypothetically defined. They include measures in the form of how well a certain task is performed, and of how hard a task can be performed with a certain degree of success. Consequently, the methods of devising and using educational measurements also vary widely—too widely for any unified exposition. What will be said about methods here will, in fact, comprise only certain recommendations and cautions which are likely to be often appropriate.

Consider first certain principles of method designed to ensure reliable measures, or at least measures whose degree of unreliability is known and can be allowed for. These are:

At least two specimens or samples should be taken of any fact about which a statement is to be made. If any individual's achievement in drawing is to be reported, use at least two drawings. If the achievement of a class in addition is to be reported, use at least two tests, preferably on two days. If the effect of a method is to be estimated, test the method with at least two classes taught by different teachers. If the quality of a specimen of handwriting is to be reported, have at least two judges rate it independently. It will often appear from the comparison of two samplings of a fact that many more samplings are needed to permit a statement that is precise enough for the purpose in view.

No fixed rules can be given, since the purpose in view determines the degree of precision that is required, but it may be noted that a test which gives, for a single pupil, an approximation so rough as to be almost useless, gives for a class of thirty-six a result

which is six times as precise, and for a group of nine classes a result which is eighteen times as precise. Ten times as large a sampling of the product in question is required to measure a single pupil as to measure the average of a hundred pupils (to the same degree of precision). In general, eight tests of 15 minutes each are superior to four tests of 30 minutes each, and still more superior to two tests of 60 minutes each, since the accidents of particular temporary circumstances are thus reduced in influence.

Consider next certain principles of method that need to be observed if we are to secure measures whose significance is certain.

Great care should be taken in deciding anything about the fate of pupils, the value of methods, the achievement of school systems and the like from the scores made in a test, unless the significance of the test has been determined from its correlations. For example, it cannot be taken for granted that a high score in checking letters or numbers is significant of a high degree of accuracy and thoroughness in general. Letter-checking tests have been so used, but with very little justification. Courtis has given reason to believe that a test with stock problems from text-books in arithmetic may be a very inadequate test of ability to reason with quantitative facts and relations, this ability being in such a test complicated by, and perhaps even swamped by, the ability to understand the verbal description of the facts and relations.

A pupil's score in a test signifies first, such and such a particular achievement, and second, *only whatever has been demonstrated by actual correlations to be implied by it.* Nothing should be taken for granted.

The significance of one *ability* (A) for another (B), is given by the correlation coefficient r_{AB} corrected for attenuation. The significance of a *particular test sampling* (A) for the ability (B) is given by the raw correlation coefficient r_{AB}. Thus, arithmetical ability itself is significant to a high degree of promise of ability with algebra and geometry, but a five-minute test in arithmetic would be much less so.

It is unfortunately the case that we do not at present know at all well the significance of any school ability or of any of the tests which we have devised as convenient means of sampling abilities. We need not blame ourselves for this: the educational measurements now in use are much better than none at all. They do excellent service, provided inferences are made with proper caution. They will do still better service in proportion as the correlations of each are determined. This work is extremely laborious, but sound method requires it.

Consider next certain principles of method designed to free measurements from certain pernicious disturbing factors, notably unfair preparation for the test, inequalities in interest and effort, and inequalities in understanding what the task is.

The best protection against unfair preparation is the provision of many alternative tasks of demonstrated equality in difficulty. This again means extremely laborious and uninteresting work, which nevertheless requires expert talent. It should be subsidized.

There is and can be no absolute assurance of equality in interest and effort. Any educational product is a product of ability conditioned by interest. All that we can do is to choose such conditions for the test as are found to reduce inequalities in interest and effort to a minimum (that is, to show high correlations with the composite of results obtained with a sampling of all conditions likely to influence interest). There is reason to believe that, when the test is taken as a part of school work, the appeal to group competition, as in "We wish to find out whether you can do as well as the sixth-grade children in Boston did," and a promise to report the results to each individual, are useful. In the case of high-school and college students a small payment in money or release from tasks, together with the promise of a full report to each individual, seems a useful method.

Inequalities in understanding what the task is, may be reduced by a preliminary trial, identical in form with the test itself, but with very easy content, and by giving special tuition to any pupil who fails in this preliminary trial. Instructions should be in simple language and should always be accompanied by at least three concrete samples of the task.

One who is eager to find imperfections can find many in present measurements of educational products. Nor is it a hard task to make constructive suggestions for improvement. An intelligent student of education could probably in a single day note a score of sure ways of improving the scales and tests which we now use. That is really child's play. The hard thing is the actual expert work of remedying the imperfection, for this involves hundreds of hours of detailed expert planning, experimenting and computing. What is needed in educational measurement is not the utterance by onlookers of criticisms and suggestions with which the men actually at work with measurements are as familiar as they are with their own names, but expert assistance in overcoming the defect.

If those who object to quantitative thinking in education will set themselves at work to understand it; if those who criticise its presuppositions and methods will do actual experimental work to improve its general logic and detailed procedure; if those who are now at work in devising and in using means of measurement will continue their work, the next decade will bring sure gains in both theory and practice. Of the gains made in the past decade, we may well be proud.

CHAPTER III

SPECIFIC USES OF MEASUREMENT IN THE SOLUTION OF SCHOOL PROBLEMS

M. E. Haggerty
Director, Bureau of Cooperative Research, University of Minnesota

In preparing this chapter on the specific uses of measurement in the solution of school problems the writer mailed to a selected group of school superintendents, most of whom were known to have used tests and scales, a questionnaire. Chiefly it was sought to learn what changes in school organization and procedure had been made as a result of such measurement.

Among 200 replies received there were 62 which reported some conscious alteration in the work of the school following the use of a standardized scale or test. In general, these changes may be grouped under six heads as follows:

1. Changes in classification of pupils
2. Changes in school organization
3. Changes in course of study
4. Changes in methods of instruction
5. Changes in time devoted to subject
6. Changes in methods of supervision

Under these same heads it is convenient to group the remedial measures described in the periodical literature, and the discussion to follow will, therefore, make use of this classification. Under each head will be given results of the questionnaire. These results are fragmentary, but serve to indicate the range of things which school officers wisely or unwisely do as a result of information derived from educational measurements. Under each head will also be given one or more detailed examples of the type of remedial work in question. No attempt is here made to give an exhaustive review of the literature. Only suggestive cases are used.

I. Classification of Pupils

1. Changes Indicated by Replies to Questionnaire

Changes in classification were of three sorts:

(*a*) The promotion and demotion of individual pupils who were found improperly classified. "We gave certain pupils double promotions." "We demoted and promoted pupils who did not fit the grade they were in." "Bright pupils were put into relief by the tests and afterwards examined; a large number were promoted thus."

(*b*) A transfer of pupils from one grade to another for particular subjects. "Pupils were transferred in reading to grades for which the tests revealed they were fitted," "When pupils reached Quality 10, Thorndike handwriting scale, in monthly tests they were promoted into advanced section, meeting three times a week," "Pupils classified in reading according to score in test: (1) those below Kansas Standard drilled in thought interpretation; (2) those who equalled Kansas median given no extra attention; (3) those who tested a grade higher allowed to drop reading for a time and work on any study they were low in."

(*c*) A general reclassification of entire school. "I classified my school below seventh grade so pupils could make up work where they were weak and take advanced work where they were strong."

2. Detailed Example

In the University of Minnesota High School 60 pupils were accepted at the opening of the school year as beginners in the freshman class. These students were required to take English and general science and had a choice of two of the following: Latin, ancient history, and mathematics. Before the first day of school they were tested, as a group, with the Trabue completion scale, an analogies test, and an omnibus mental test. Later they were tested with Thorndike's reading scale Alpha 2, a series of arithmetical problems, and a wide range of mental and educational tests. The initial tests were given in order to get a measure of the probable success of the several students and to find an intelligible basis for sectioning the group for purposes of instruction. On the basis of the test scores, the students were divided into two or more classes for the several school subjects.

The school marks for these students were given in all by seven instructors; each student was rated by four different teachers. The marks were in letters *A, B, C, D* and *F* and were based upon a relative marking system. In the total marks to be here considered *A* was given to 8.6 percent of the class, *B* to 21.2 percent, *C* to 43.3 percent, *D* to 20 percent and *F* to 6.8 percent. In the work in general science, where two sections were taught by different teachers; the instructors conferred in giving the marks, so that the sixty pupils were rated as a single group. This rating was in large measure on the basis of objective tests which were the same for all. Similar objective tests were used in all subjects.

The marks to be here considered are those given at the end of the first month. These marks, more than those to be given later, are liable to inaccuracy as indicators of ability to do the high-school work. Now, comparing these marks[1] with the scores in the standard tests, what do we find?

The first and simplest measure of prophecy which we wish to find is the median attainment, for sixty pupils can be fairly divided into two sections for high-school instruction. If, by dividing them on the basis of the tests, we get the more capable half in the upper section and the less capable half in the lower section, the prophecy of the tests is confirmed. Figuring the median attainment, we find that the omnibus test placed 72 percent of the sixty pupils in that half of the class in which they were placed by the teachers' marks for the first month. Likewise, the reading scale Alpha 2 placed 72 percent correctly. When the scores of the two tests are combined the median retention is 76 percent. In other words, the tests placed more than seven pupils in every ten correctly before they had done a single day of high-school work.

As a matter of fact, the prognosis was even better than this, for all of the instructors were agreed that the school marks did not represent the abilities of all of the pupils accurately. Thus, one boy of acknowledged ability stood 12 in the omnibus test and 22 in the reading test. In marks he stood 57, scoring *C* in history and failing in general science, mathematics and English. Every one, in-

[1]These data are available through the courtesy of Dr. W. S. Miller, Principal of the University of Minnesota High School.

cluding the boy himself, admits that through this month he idled away his time. There is also the opposite type of case—a student scored low in the tests but by unusual industry lifted himself above the median line in marks. When one considers the part which industry plays and the variable factors of outside work, social distractions, parental interests and personal attitudes in relation to school achievement, it is clear that intellectual ability alone is not determinative.

It is encouraging to believe, however, that approximately eight children in ten are properly placed by such tests. This idea is supported by figuring the average mark of the students in the upper and lower halves of the tests. Equating the letters to the figures 5, 4, 3, 2 and 1, it is evident that a student carrying four subjects could score 20 points. As a matter of fact the upper half as divided by reading test scored 13.4 points, while the lower half scored 10.3, and similar figures for the two groups as divided by the omnibus test are 13.5 and 10.2. If we divide the sixty pupils into three equal groups on the basis of the two tests, the upper third averages 14.2 in marks, with no failures in a total of eighty marks. The lower third averages 9.8 with 17 failures in a total of eighty marks and only 9 marks above *C*.

Reasoning from all these data, it is safe to say that both educational and mental tests may be used to classify students for purposes of instruction; that a few hours spent in preliminary examination will foreshadow later achievements to a high degree. Admittedly, much technique remains to be developed, but the promise is sure. Superior students can be detected and grouped together; mediocre students can be put with mediocre students, and weak students, instead of being submerged in the struggle to maintain standing, can receive the help they need. It would be difficult to overestimate the increase of efficiency that would come from the better adaptation of instruction in consequence of such classification.

How dependable the tests really are in promoting superior students is evidenced by numerous instances, of which the following may serve as an example:

Robert was 12 years old, beginning second semester of eighth grade. His teachers reported him indifferent, doing only ordinary work and inclined to be the center of schoolroom disorder and organized insurrection. Parents noted that, though previously much interested in school, the boy now disliked to attend; he disliked the teachers and wanted to drop out. Robert insisted that the studies were not interesting, that he knew all he wanted to know about them already. Mental examination showed an intelligence quotient of 142, a mental age probably greater than that of some of his teachers, who bored him to death by treating him as an ordinary twelve-year-old. He was recommended to high school, entered three weeks late, led his class at the end of six weeks and at every subsequent interval when marks were given. More important, his whole attitude toward school was changed, because the advanced work was a real challenge to his mental ability.

II. School Organization

1. Changes Indicated by Replies to Questionnaire

Seven types of change in school organization were indicated as follows:

(*a*) Change in size of classes. "Smaller classes in arithmetic." "More teachers in arithmetic." "The services of additional teachers demanded for defectives in industrial training." "Enlargement of class for defectives."

(*b*) Division of classes into special sections. "An advanced and a special section made in writing on the basis of errors."

(*c*) Organization of Special Classes. "Opened special room for backward pupils." "Organized special corrective work."

(*d*) Departmentalization. "Departmentalization of sixth, seventh, and eight grades." "Department teaching."

(*e*) Arrangement of parallel programs. "Spelling periods for different grades arranged for same time to permit transfer of brighter pupils." "Same for reading."

(*f*) Appointment of supervisors and supervising principal. "Position of supervising principal for primary grades created." "Appointed a director for a newly organized Bureau of Research." "Plan to hire a trained supervisor for writing next year."

(*g*) Inauguration of special schools. "An initial attempt to develop an elementary industrial school for pupils shown by the tests to be unfitted for the regular work."

3. Detailed Examples

(*a*) *Parallel Programs.* At the Lake Harriet School in Minneapolis a measurement by the Ayres scale showed a wide distribution of spelling attainment among the pupils of each class. In the 8A grade, 43.3 percent of the children were of eighth-grade spelling ability. Grouped along with them were 26.7 percent of seventh-grade ability, 26.7 percent of sixth-grade ability and 3.3 percent of fourth-grade ability. More variable than this was the fifth-grade class, where the distribution showed every level of ability from the second to the eighth grade. On the basis of this showing, Miss Probst, the principal, observed that it was practically impossible in group instruction "to devise a spelling lesson which would tax the capacity of each individual in the group." It was determined, therefore, to rearrange the grouping in such manner as to give every child "capacity work." To justify this regrading, further tests were given. With the results of all the tests as a basis, the pupils were redistributed so that those of like ability recited together. In the new organization the original 7B class retained four pupils from an original total of thirty-two. The others were distributed to new groups as follows: seven to 8A, six to 8B, thirteen to 7A and four to the sixth grade. To the new 8B group, the several grades contributed in this fashion: 8A gave 7 pupils; 8B gave 14; 7A gave 4; 7B gave 6; the sixth grade gave 6; the fifth grade gave 5, and the fourth grade one—a total of 42.

When the redistribution of the pupils on the basis of ability had been made, the spelling recitations for all were arranged to take place between 11:45 and 12:00 o'clock on each day. At the sound of the bell, each pupil passed to the room where his particular level of work was in progress. There he would go each day until a definite change in his work occurred. If the average of his work for two consecutive weeks should fall below 90 percent, he automatically gravitated to the grade below. If he maintained a standard of 99 percent for four successive weeks, he was promoted to the grade above. To illustrate: two 8B boys whose beginning record showed about 65 percent dropped back to the fifth grade. Unable to maintain the pace of this grade, they dropped back to the fourth. "Here," to quote Miss Probst, "they evidently found their spell-

ing level, for they took a fresh start, were promoted to the fifth grade, then to the sixth and finally to the 7B group."

The chief outstanding result of the experiment was a very great diminution of the variability in the several grades. Whereas in the first test but 43.3 percent of the 8A pupils had eighth-grade spelling ability or better, 71.4 percent came up to this standard in the second test in May. In the 7A class, the corresponding figures were 45 and 80: in the sixth grade they were 55 and 90.

(*b*) *Special Classes in Schools.* A second change in school organization brought about by means of educational and mental tests is the institution of special classes for the education of exceptional children. Wallin reports (1914) eleven types of such classes.

(*c*) *Organization of Bureaus of Research.* Probably the most significant change in school organization growing out of the measurement movement is the organization of Bureaus of Research as a supplementary supervisory and administrative agency. Since this topic is treated at length in other chapters of this report an extended statement is unnecessary at this place.

III. Course of Study

1. Changes Indicated by Replies to Questionnaire

It was difficult often to differentiate changes in the course of study from changes in methods of instruction. Under this head four types of change were more frequently noted than any others.

(*a*) Change of Textbooks. In some cases the textbook was merely changed for another book. In other cases, the book was apparently dropped altogether and other types of material substituted.

(*b*) Emphasis was changed by giving more space in the course to different parts of a subject. This seemed to be especially true in arithmetic and in spelling.

(*c*) Numerous replies indicated that the tests served to fix standards of achievement for different grades.

(*d*) The organization of specialized curricula for special classes.

2. Detailed Examples

(*a*) *Specific Aims to be Achieved.* The effect of educational measurements on the course of study has been to fix specific aims in the several school subjects. With most of the standard scales and tests there have been proposed ideal forms for the several grades, or the average and median scores made by these grades have been set as desirable ends to be achieved.

The specific aims to be accomplished may be grouped under three heads: aims in rate of work; aims in quality of work; and a combination of the two, under the head of efficiency.

The accompanying table contains a representative set of such aims. The full meaning of these standard scores will be understood only in connection with the tests themselves.

SAMPLES OF SPECIFIC AIMS

ACTIVITY	MEASURED BY	SCORE
Rate of Silent Reading	Gray (Ancient Ships)	2.87 words per second
Quality of Silent Reading	Thorndike Scale for Understanding of Sentences	7.50 on Scale Alpha 2
	Word Knowledge	8.50 approximately on Scale A, A2 or B
Addition	Courtis Series B	Rate 12 Accuracy 100 percent
Addition	Woody Series A	Accuracy 9.01 scale points
Reasoning	Stone	Rate 8.75 Accuracy 90 percent
Spelling	Ayres scale	100 to 50 percent on Columns N to Z
Writing	Ayres	Rate, 79 letters per second Quality, 62 Gettysburg scale
Language	Trabue	14 to 16 scale points
Composition	Harvard-Newton	70 on scale
Grammar	Buckingham	5.13 questions answered
Geography	Hahn-Lackey	0 to 99 percent on 20 sets of questions

Any pupil who at the middle of his eighth school year can achieve the above standards should be considered normal for the grade. A lower attainment clearly indicates the need of remedial work for the class or individual in question. A distinctly superior achievement is evidence of superior intelligence on the part of the pupils or of superior methods of instruction.

(*b*) *Minimal Essentials.* A second important movement making use of measuring methods is the effort to derive minimal essentials. Of the nine methods of determining the minimal content of the course of study described by Coffman, at least four are essentially methods of measuring the acquired behavior of children and adults; two are measurements of the content of reading matter, and the other three are concerned with pooled opinions. Of the ten chapters in the *Sixteenth Yearbook* of this Society, "Minimal Essentials in Elementary School Subjects," five make direct use of information derived from the measurement of children's attainments. The other five are based on measurements of the content of books and other published material.

It is unnecessary to enter here into a detailed statement of these investigations. It is sufficient to note that the determination of the content of the course of study must follow two fundamental principles suggested by the following questions: (1) What should children know and do as children and adults? (2) What can children at any age learn with profit? The final answer to both of these questions must be obtained by measurement.

IV. Methods of Instruction

1. Returns from Questionnaire

(*a*) *Increased Emphasis.* Approximately one-third of the replies note "increased emphasis," which apparently means more time given to a subject or more value placed upon efficiency in it. The following are representative of these replies:

"More emphasis, all grades, on meanings of words and sentences." "Stressing legibility in writing." "Greater emphasis on fundamentals in arithmetic." "More emphasis on correct use of words in reading, less on definition." "Special emphasis given to those subjects where standard was low." "More drill in all grades." "More silent reading."

(*b*) *Drill.* The favorite recourse for improving attainments which the tests show to be poor is "drill." This term is vague, often meaning merely more time devoted to repetition of certain processes.

"Arithmetic: All teachers required to give more drill work." "Special emphasis given to those subjects where standard was low." "Five-minute daily drill on fundamentals in arithmetic." "More time and attention given to drill in number combinations." "Dictation drills with attention on punctuation." "More intensive drill in grammar and punctuation." "More drill in spelling."

(*c*) *Specialized Drill.* Some correspondents report a particular type of drill, involving a definite change in the details of the drill process as well as change of time

"Courtis drill cards in arithmetic in two rooms; more oral drill in all." "Horace Mann method of spelling adopted." "Installed Palmer method of writing."

(*d*) *New Devices.* Another method of improvement of instruction was the invention of special devices.

"Made room charts showing individual's work (median, quartile, safety zone)." "Teachers used questions similar to Kelley test, and applied to geography and other subjects." "More instruction through interest of pupil." "Supervised study periods—all grades." "Three periods of supervised study added to school day: those who failed in one subject required to stay 1 period; in two, 2; etc." "Greater use of dictionary for meaning of words." "Tests devised to watch pupils' progress."

(*e*) *Individualized Teaching.* The measurements serve to fix attention on individual differences among children and to further the adaptation of instruction to individual needs. Illustrations:

"Individual attention; specifics devised for securing appreciation of good writing." "Individual help given slow pupils." "Methods adapted to ability of pupils." "Backward pupils discovered and given special attention." "Promotion more by subjects."

2. Detailed Examples

(*a*) *Methods of Drill in Arithmetic.* Mead, at the request of Superintendent Condon of the Cincinnati schools, undertook the experimental evaluation of two kinds of practice material in the fundamentals of arithmetic. The materials in question were the Courtis Standard Practice Tests and the Thompson Minimum Essentials in the four fundamentals. About 900 fifth-grade pupils

from fourteen schools were divided into two approximately equal groups and were drilled 15 minutes daily from February to May. The efficacy of the two kinds of practice material was determined by "preliminary and final standard tests with each group" (Courtis Series B). The initial and final tests, as well as the intervening practical exercises were given by the class teachers under standardized instructions, and care was taken to keep conditions constant and favorable.

The results of this carefully controlled experiment show that in *speed of work* the two kinds of practice material produce improvement, and essentially the same amount of improvement. Both also show gains in *accuracy of work*, but they differ essentially in the amount of improvement in accuracy resulting from the exercises. The pupils of the twelve classes using the Thompson Minimum Essentials show median gains in percent of accuracy for the four fundamentals of 2.5, 4.0, 2.9, and 15.7; similar figures for the Courtis Practice Tests were 9.7, 8.1, 8.9, and 18.2. Clearly, the Courtis Practice Tests are superior, and the test shows in what respect they are superior.

It does not follow from this that the Thompson Minimum Essentials are of no value or that in another grade and under other conditions they might not prove superior. What is fairly certain is that, upon the use of the two methods in fifth-grade classes under school conditions prevailing in Cincinnati schools from February to May, 1916, and described by Mead, the debate has closed.

(*b*) *Teaching of Handwriting.* The "two-squad" method has been used by Mr. A. G. Capps in measuring the effect of "diagnosis and corrective measures in the teaching of handwriting." The handwriting of 44 sixth-grade children was measured on October 9. All who scored high on the Thorndike scale, Quality 9 or more, and whose papers were relatively free from common errors, were put into an "advanced class." All others were put into a "special class." The advanced class was taught three days per week; the special class, five days per week. A pupil was allowed to pass from the "special" to the "advanced" class, when on the monthly tests he scored Quality 10 (Thorndike).

All pupils were treated by the same general method, which was as follows: The handwriting of each pupil was diagnosed following Freeman's analysis of "quality" into alignment, spacing, slant, quality of line, and letter form. On the basis of this diagnosis, remedial treatment was prescribed for each class and for each individual, and detailed methods of instruction were worked out. The pupils were made acquainted with their own difficulties and taught to practice with a view to achieving certain detailed aims, such as better form for the letter *a*, improved alignment, etc. The experiment was continued for five months, and standard tests were given at four-week intervals.

The significant results of the experiment may be stated as follows: (1) The advanced class improved in quality from 9.06, or less than fifth-grade attainment, to 10.47, or slightly more than seventh-grade scores (60 minutes per week). (2) The special class improved from 8.74, or fourth-grade score, to 9.68, or slightly less than sixth-grade average (100 minutes per week). In both classes there was a substantial gain in speed.

In the light of these results, regarded as tentative by the experimenter, there can be little doubt this teaching method has considerable claim to efficiency.

(*c*) *Improvement in Written Composition.* Somewhat different in type is the study by Brown and Haggerty of the improvements in English composition. The weekly composition exercises of three high-school classes through a period of twelve weeks were measured by the Harvard-Newton Scale. By this method it was possible to secure a "learning curve" for individuals as well as classes, and thus to see the educational behavior of a student somewhat more intimately than when only initial and final measurements are made.

From this experiment several facts stand out. (1) There is no essential difference in composition performance among classes rated as first-semester freshmen, second-semester freshmen, and first-semester sophomores. (2) In general, classes gain in power to write during a twelve-weeks period about four or five Harvard-Newton Scale points. (3) Some students, relatively poor ones, gain much more than the class average, in some cases as much as

twenty scale points. (4) Other students, often the superior individuals, make little or no improvement; some individuals do more poorly at the end of the twelve-weeks period that at the beginning. (5) Some composition topics elicit better products than others, as shown by the median scores of all classes.

This simple experiment does not get very far in solving the manifold perplexing problems of composition teaching. The method, however, offers a means of further investigation.

V. Time Distribution

The changes in time were mostly of the nature of increased time to subjects where the tests showed the product to be low grade. The following are characteristic replies:

"Lengthened periods for subjects where deficiency was greatest." "More time given to arithmetic." "More time given to writing in all grades." "Ten minutes of arithmetic recitation period used for drill three days per week." "Increased number of hours for industrial work."

Only one correspondent reported a diminution of time, saying, "We excused pupils and grades from work 'over done'." Numerous studies, from that of Dr. Rice on, have shown that results of school work are not directly correlated with the time spent on the subject. Studies in handwriting, spelling and arithmetic show that the maximal time may yield poorer returns than a smaller amount of time.[2]

VI. Changes in Supervision

As a rule, tests have been introduced into the schools through the supervisory and administrative officers. It is not surprising, therefore, to find that the results of the tests have had a direct and considerable effect on the detailed work of such persons.

The general supervisory practice is to report the results of tests to principals and teachers. Such a report gives not only the scores of the class or classes concerned but also comparable norms from other classes within the system and from other systems. Apparently some supervisors end their use of the tests with this report.

[2]Some concrete studies in this field were reviewed by the writer in *School and Society*, November 19, 1916.

A second step taken by other supervisors is a conference between the teacher and supervisor, a conference in which "the opinion of the supervisor is not pitted against that of the teacher, but in which the attention of both is directed to the definite and objective result of the test and to the causes producing such result." Growing out of such conferences come "new outlines of work," "re-organization of programs," "special lessons for teachers in teaching spelling," "requirements in the handwriting of teachers," in fact, any one of the changes hitherto enumerated.

A third step to be noted is a further examination of a class or pupil for the extension of the diagnosis. This is made with other tests and through personal investigation.

A fourth step taken by some supervisors and teachers is a second measurement, after a period of remedial work, to test the efficacy of the changed program. In about one-fifth of the cases where remedial work was reported, the supervisor had used the tests in this way. The following are typical replies:

In arithmetic "Graph showed greater improvement in grade in one month, than first graph showed from class to class (five months' work)." "Have used drill and are now *above* the Courtis Standard." "The weaker pupils do better work. More pupils brought up to required standard. Pupils more accurate." "Plateau disappeared which had existed from Grade VI to VIII; curve for both attempts and rights in all operations shows gradual development to Grade VIII."

In reading: "Improvement in median score from September 7 to January 23; Grade VII, 5.3, Grade VIII, 5.6."

In spelling: "Pupils average a grade higher." "More uniformity of grades and children know where they stand."

In writing: "Better quality. Both quality and speed nearer average for grade."

A fifth important use of tests by supervisors is to keep track of the normal progress of students and classes. Results of such tests serve as a sort of weather map of the school system and show the prevailing winds. They enable the supervisor to know in an intimate way the entire system and to direct his efforts at supervision where they are most needed. Where fair progress is being made he can give little attention and save his usually inadequate energy for the places where it is most needed.

Finally, it is apparent that supervisors are making the class attainment, as shown by the results of measurement, one factor in the rating of teachers and for the further professional training of teachers. The following replies show this:

"Introduced a system of teacher training." "Certain teachers required to go to summer-school." "Teachers required to visit classes with strong teachers." "In second semester, changed teacher whose class was weak to a lower grade class, to test power and presentation of subject." "Certain changes in both principals and teachers, and will require certain teachers to attend summer-school."

The experience of one supervisor is related in a most interesting paper on the Supervisor's Use of Standard Tests, by J. C. Morrison, who presents the details of a year's work in Chatham, Mass., and sums up the work in these words:

"Standard tests have proved an effective means in supervision. Through their use teachers are improving the technique of their method and exercising a nicer judgment in the relation of subject matter. Pupils are working to exceed their own record. Teachers and older pupils are coming to understand the scientific idea of education as it applies to the ordinary classroom. In no other way could the principal gain so close a knowledge of each individual child in the school. This knowledge is serviceable in placing new pupils, in determining promotions, in selecting accelerates and defectives, in searching out the special difficulties of the individual, and in gaining the cooperation of the parent. The results of the tests have proved of interest to the public. On the basis of these results the board of education has employed a teacher to give approximately one third of her time to the testing and supervision of work with special pupils. The school has made a start in the study of its children and will eliminate a large part of the wasted time and effort that results from the choice of the wrong high-school course."

In this survey of remedial measures based on the use of standard tests one observes that many of the changes made are the time-honored ones which school officers have traditionally made on the basis of personal opinion and in response to changing ideas. Apparently what the tests do in such cases, is to render definite the arguments for these changes and to make accurate the evaluation of the efficiency of remedial measures when once they have been car-

ried out. To a schoolman who cares to guide his practice by facts rather than by debate, this service sufficiently justifies the idea of measurement. To argue the case with other types of individuals is, perhaps, a waste of time.

One thing the tests show, however, which traditional practice has never recognized, because it never knew it, and that is, the enormous range of individual differences among children, both in ability and in attainment. These differences the tests reveal in a way that must inevitably alter profoundly our whole program of education.

CHAPTER IV

GENERAL ORGANIZATION OF EDUCATIONAL MEASUREMENT WORK IN CITY SCHOOL SYSTEMS

FRANK W. BALLOU
Director, Department Educational Investigation and Measurement,
Boston, Massachusetts

Every change which takes place in educational practice is preceded by a period of agitation. The profession must be made aware that a given condition is unsatisfactory and must be convinced that the proposed change will bring about improvement. In addition, the lay public must be educated to understand the meaning and significance of the proposed change. Many educational schemes never proceed beyond this stage of professional and public agitation.

Following the period of agitation comes a period of trial and experimentation, in which the proposed change is subjected to close scrutiny. Generally, both the profession and the public produce searching and frequently unjust critics. These criticisms usually prevent even a trial of any changes in our educational practice which do not give promise of a reasonable degree of success.

In those cities where special research departments have been organized, educational measurement is established beyond the stage of argument or debate. There the movement has passed successfully through the periods of agitation and of experimentation. Even through the country generally, the movement has received such general recognition and endorsement during the past few years that progressive school systems do not now need to be provided with arguments in favor of educational measurement. This chapter, therefore, is not an argument for the introduction of educational measurement: it is not a discussion of the debatable phases of the subject; neither is it a delineation of the educational advantages to be secured by measurement. Rather, it is a description of ways

and means of introducing systematic educational measurement in a city school system and carrying it on successfully. The methods suggested are those which appear to have succeeded best in those city school systems that have undertaken to measure educational results in an organized way.

The time for the introduction of educational measurement into a school system should be wisely chosen; standard tests should be given only after the way for them has been carefully prepared. The success of any educational reform depends on the intelligent cooperation of the members of the educational profession. This is especially true of educational measurement, because it involves an entirely new method of attacking educational problems and an altered attitude on the part of teachers toward the whole educative process. Success is contingent on a thorough understanding of the aims, methods, possibilities, limitations, and achieved results of the use of standard tests and scales. Unless and until the profession is so informed, and as a result is prepared to cooperate effectively in its administration, the possible values of educational measurement are not likely to be secured. Proper instruction through lectures, talks, and teachers' meetings will do much toward preparing the way for success in carrying on such work.

Educational measurement involves selection of the tests to be used, testing the children, marking or scoring the papers, tabulating and interpreting the results, and utilizing the conclusions as a basis for securing improvements in teaching when the outcome proves that the present teaching is unsatisfactory. This chapter deals with each of these topics.

A. Selecting the Tests

The attitude of teachers and others towards educational measurement depends largely on first impressions. The importance of making favorable first impressions, therefore, cannot be over-emphasized. The first standard test selected to be given in a school system should be one which is most likely to be favorably received by teachers. Fairly satisfactory tests for several subjects or phases of subjects are now to be had.[2]

[2]See Chapter VII for a list of available standard tests and scales.

Adequate standard tests for any school system should possess the following characteristics: they should measure educational products obviously within the scope of the course of study of the city; they should aim to measure those subjects or phases of subjects which are clearly measurable; they should be reasonably simple; they should be accompanied by adequate instructions as to how they are to be given and how the results are to be scored; they should be scored and tabulated with reasonable ease; they should already have been given to a sufficient number of children so that well-founded standards of achievement have been established.

The demand for standard tests in various subjects has been so great that some questionable tests have been put on the market. For example: a so-called 'standard test in spelling' is available which tests the ability of children in Grades V to VIII to spell such words as the following: *nunciature, sphericity, hoggery, senescent, symmetrize, incremental, rigmarole, verisimilitude, anthropometric, tubule, erosible,* and *divestiture.* These words are not to be found in the course of study in most school systems in the United States. Lack of ability to spell such words cannot be charged against the schools, because the schools have not undertaken to teach children to spell them. This test and others like it that do not measure classroom instruction, should be avoided—at least in the beginning of testing work. If they are used, the purpose in giving them and their limitations should be distinctly understood. There are satisfactory standard tests in spelling which are obviously within the scope of classroom instruction.

Criticism has been urged against educational measurement on the general ground that important products of good teaching in certain phases of all subjects and in all phases of certain subjects cannot be quantitatively measured. Whether this is so or not may be debatable: let the debate go on. It is not necessary, however, to postpone all educational measurement until the debate is settled. In the meantime, giving standard tests in measurable subjects or in measurable phases of subjects will be profitable for teachers and pupils, and may also furnish valuable information by which to determine to what extent the results of educational practice generally are measurable.

In introducing educational measurement into a school system for the first time, the more simple the tests are, the better. To be usable, standard tests must be accompanied by adequate instructions for giving them and for marking the papers. Unfortunately there is no royal road to travel in carrying on educational measurement. The details of giving the tests must be mastered by those who would make proper use of them. If tests are worth giving at all, they should be given according to the systematic plan which the author of the tests has worked out. Success in introducing testing into a school system will be more certain if the tests are selected only after careful consideration of the instructions for giving and scoring them.

It is also important that well-recognized tests be used in the initial testing in a school system. Not everyone who desires to give standard tests has the time, resources, or qualifications for preparing his own tests. It cannot be too urgently recommended that at first tests should be selected from those already available. Later, one may successfully experiment with the preparation of the tests. Among the gravest dangers which educational measurement faces today is that which arises from amateurish attempts at the construction of standard tests by those who do not realize the need for a careful testing of the tests themselves before they are published for general use.

Well-recognized tests are urged for use; first, because such tests are undoubtedly superior to those which would be prepared under most circumstances by a beginner; second, and more particularly, because the results achieved in any school system can thereby be compared with similar results from other school systems.

One of the desirable outcomes of the giving of standard tests is the establishment of standards of achievement. Such standards furnish measures with which one may compare his own results. Standards are adequate, however, only when they are based on a large number of results. Tests which have been standardized from the results achieved by a small group of pupils in one or two school systems are to be avoided unless there are other quite special reasons for using them.

Testing should always be purposeful. Testing children merely for the sake of giving tests cannot be too thoroughly condemned. The following principle of practice is commended: never give a standard test unless you have a definite purpose related to the improvement of instruction and unless you are prepared to tabulate, interpret, and *use* the results at once.

Since the ultimate purpose of all educational measurement is the improvement of the instruction, it also becomes of paramount importance that the results be made known to the teachers and others before their interest in the tests has waned.

B. Getting Tests Given

Let us repeat that if the results obtained from giving standard tests are to be worth the time, money, and effort expended, the tests must be properly given. This means not only that they must be given in the same manner in every classroom throughout a school system, but it means, also, that they must be given according to the directions that accompany the tests. If this is not done, the results cannot be compared with the standardized achievements in other cities—a fact which renders the interpretation of the results very difficult, if not impossible.

It may be assumed that the author of the tests had reasons for adopting the procedure indicated by his instructions, and that the most effective use cannot be made of the results obtained in any school system unless those directions are followed. Even though one does not agree with the directions for conducting the tests, one would better not use the tests at all than to deviate in any important respect from the author's directions for giving them.

Tests may be given either by principals, supervisors, or teachers, or by persons especially trained for the purpose. Such tests as spelling tests, which are among the simplest to give, may be given satisfactorily by principals or teachers. They do not require careful timing, and the directions for giving are comparatively simple and are easily followed. With a minimum of instruction any principal or teacher can successfully conduct spelling tests.

Such tests as the Courtis standard tests in arithmetic offer a more difficult problem. They must be accurately timed and the

conditions under which they are given must be controlled. Those who give such tests must be specifically trained for the purpose.

For this purpose one of two methods may be followed: a group of teachers or principals in the service may be instructed, outside of school hours, and assigned to give these tests in the various classes throughout a city. Or, as in Boston and other cities, those preparing to become teachers may be trained to give these tests as a part of their preparation for teaching. In Boston each member of the senior class in the Normal School spends one month in the Department of Educational Investigation and Measurement and receives instruction in the meaning of educational measurement and the training necessary for giving such tests as the department desires to have given. Following this training, these seniors are assigned to give the tests in the various schools of the city.[2] Very similar arrangements are made in many other places. Cities in which a college or university is located have also used trained college students of education as special examiners.

Experience has not yet determined which practice is likely to be followed in the future. It is certain that whoever gives tests must receive proper instruction in the methods of giving them. The policy of training normal-school seniors in measurement work is based on the assumption that the teaching staff of a city ought to be competent to give such tests as are needed for the measurement of the work of a school system. If the teachers of a city can be so trained, this is undoubtedly the cheapest and most effective method of solving the problem. If, for one reason or another, teachers do not prove satisfactory as examiners, then it necessarily follows that competent specialists must be employed for giving the tests, as well as for carrying on other phases of educational measurement. Whatever be the method finally adopted, those who give tests must not only be properly instructed, but their work must also be adequately supervised. This supervision is the function of a department of educational research. In the absence of such a department in a school system, the superintendent's office should assume this responsibility.

One of the difficulties in relying on the teaching staff as a whole to give standard tests grows out of the attitude of teachers

[2] See article by the writer in *School and Society*, Vol. V. pp. 61-70.

toward such tests. In their attitude toward educational measurement one may expect to find three distinct classes of teachers in every school system: (1) those who endorse it, (2) those who are indifferent to it, (3) those who oppose it. On this account it is probably wise to introduce standard tests into a school system on a voluntary basis; that is, to give the tests only in those schools or classes where principals and teachers are willing to have them given.

The success of educational measurement in a school system depends on doing a small amount of testing and *doing it well,* rather than on doing a large amount. Quality rather than quantity is more likely to win favor. Opposition to educational measurement can best be disarmed by showing the improvement secured from the results of standard tests.

C. Scoring the Papers

The amount of time involved in giving tests is small compared with the time necessary for correcting the papers. Assuming that the teachers give the tests, who shall mark them? The answer to this question depends largely on the character of the tests. Certain types of papers from some tests can be satisfactorily marked by the pupils under the supervision of the teacher. For example: if a spelling test is given by the teacher, she may at once spell the words aloud and have the children mark the errors, following the common practice. The pupils' scoring in every case, however, should be properly checked by the teacher. Likewise, in some tests in arithmetic, children may be provided with answer cards and shown how to correct the papers.

The scoring of many tests, however, involves the exercise of some judgment, and pupils cannot be depended on to score such tests. In most cases, papers in geography and history should not be scored by pupils. Further, if one desires to give credit for a correct method in problem work in arithmetic, even though the answer is incorrect, the judgment of the teacher is undoubtedly necessary.

Teachers should not be burdened with an unreasonable amount of work in the giving of standard tests. A satisfactory principle to follow may be stated as follows: teachers should be expected to correct test papers and tabulate results only in so far as this work

will increase the teacher's knowledge of the abilities of pupils in her class. Putting this principle into practice would mean that in most cases the teacher may be expected to mark the individual papers of pupils in her class. She may be expected, also, to make class tabulations or summaries of individual achievements, because only by so doing can the teacher learn where her class stands in relation to other classes in the school, or in relation to the general standard of achievement established for the test. Beyond class summaries the work is likely to be too largely clerical to be of direct value to the teacher, and should be carried on by others.

In marking papers and making class tabulations, teachers should have some one to whom to look for instruction and guidance. If there is no regularly organized department to supervise and direct educational measurement, some qualified person in the system should be assigned to do it. Through teachers' meetings held after school, for which classes may be dismissed a half hour or an hour early and through instructions issued from time to time, the supervisor of educational measurement can materially lighten the burden which otherwise falls on teachers in this type of work.

D. Tabulation and Interpreting the Results

In Boston the Department of Educational Measurement has found an effective and economical way of making grade summaries, school summaries, and city-wide summaries of testing results. Girls from the Boston Clerical School are assigned to this work. These girls are being trained specifically for clerical work and on graduation from school may take positions which require skill in just the kind of work involved in making tabulations from tests. The Department asks the Clerical School to send relays of eight or ten girls as long as the work lasts. Each group of girls reports for three or four days to the office of the Department where the tabulating is to be done. The Department instructs them in the special methods of tabulation, and has the work so systematized that a definite record is kept of the speed, accuracy and effectiveness of the work of each girl. Their attendance is also kept and reports are made to the head master of the school when the girls return. The teachers in the Clerical School consider this work an important part of the practice and

training which the school desires to give. The Department has found it a satisfactory method of getting the work done. The expense to the city involves only the car-fare of the students to and from their homes to the office of the Department. Much the same method has been followed in other cities that have commercial departments in their high schools. It affords a practical solution of what is often a critical problem in measurement work.

If teachers or prospective teachers are trained to give the tests, and to score the papers of their respective classes, and if commercial-high-school pupils or normal-school pupils are used to make grade summaries, school summaries, and city-wide summaries, little or no direct expense is involved in that part of the work. But even though members of the teaching staff or pupils in the school system are thus employed, many additional tabulations will be desirable and necessary if the results are to serve their greatest usefulness. For such work it is essential that competent clerical help be provided. It is the kind of work which is not easily done by those unfamiliar with it, and is likely to be tedious as well as voluminous. To be done effectively it should be done by those who understand it and who have more than ordinary interest in it. Such persons can be found in every school system and when found, should be assigned to render this kind of service. No city should undertake educational measurement without understanding that it involves some expenditure: a portion of this should be devoted to the securing of competent clerical assistance for statistical tabulations and another portion to the general direction and supervision of the work (by the superintendent of schools in the smaller cities and by a special school official in the larger cities).

After the various tabulations and summaries have been made by grades, by schools, and for the city as a whole, the results must be interpreted. This is one of the most important phases of educational measurement. On the interpretation of the results really depends the usefulness of the tests. One of the greatest dangers involved in educational measurement throughout the country today is the fact that many persons are giving tests who are not competent to tabulate and interpret the results. In general, interpretation should be made by the superintendent, or by some other school

official thoroughly conversant with all the local conditions. The best results will be obtained when some competent person, who is especially interested in and trained for such work, is engaged by the board of education to supervise the giving of tests and to interpret the results to the teaching staff.

E. Making Use of the Results

Successful educational measurement work in a school system involves not only selecting the best available tests, giving them according to directions, scoring the papers, tabulating the results, (all at a minimum of cost in time, energy, and money) and interpreting the results, but it also involves getting the information derived from the tests to the persons concerned, getting those persons to consider it, and getting them to do something about the conditions revealed, if those conditions demand it.

Obviously, some sort of report must be prepared by the official who interprets the results of the tests. Typewritten or mimeographed copies of the report may be made for a small school system, but if a large number of copies is needed, the report should be printed. If printed, it is more likely to be given consideration by those concerned and to be kept for future reference and guidance.

Much of the success in getting the desired information to those concerned will depend on the character of the report. With conditions as they are at present, the facts must as far as possible be stated in simple, non-technical language. Few teachers now in the service have any knowledge of the technical terms employed in educational statistics. If such terms are used, they must be carefully defined. Our courses for the training of teachers are now beginning to include some study of educational measurement. It will be a generation, however, before the teaching profession as a whole can be presumed to have had training in educational statistics. Until that time it will be necessary to prepare non-technical reports for the teaching profession.

Finally, it should be emphasized that standard tests are for diagnostic purposes: they show the abilities and needs of children. Recently one teacher said: "My class has taken the Courtis tests in arithmetic twice each year for three years, and I do not see that it has done them any good!" On investigation, it was found that this teacher had made no effort whatsoever to use the information furn-

ished by the tests as a basis for correcting the many weaknesses shown among her pupils. There is no special magic in standard tests which will work any educational miracles on pupils who take them. The teacher must be made to realize that the results from the tests are for her information. Unless she is shown how to make use of them, and is willing to do so, standard testing *is not worth while*. On the other hand, if the teacher is willing and intelligent, the information derived from standard tests is of the greatest service. For it enables her to reach desired goals by the most direct path, with the least expenditure of energy and labor on her part and with the greatest benefit to the children.

CHAPTER V

BUREAUS OF RESEARCH IN CITY SCHOOL SYSTEMS

EUGENE A. NIFENECKER
Assistant Director, Bureau of Research and Reference, New York City

The city bureau of educational research is the direct and logical outcome of the combination of the survey movement and the movement for the use of measurement. A careful appraisal by a group of experts from without has been demanded and paid for by community after community. Very evidently such stock-taking is regarded by the public as beneficial in its effects. Too often, however, the result, as far as the local schoolmen are concerned, has been disastrous. The time spent by the experts in the local field is necessarily short, their knowledge of local conditions necessarily limited and their interpretation of results has sometimes seemed to local school officials unwarranted and unjust. More often than not, superintendents have attempted to explain away the defects revealed and to make light both of the facts and of the recommendations of the survey experts. But facts are stubborn things, and changes of superintendents and upheavals and reorganizations of various sorts have inevitably followed the adoption of such a policy. So it has come about that many times a survey report has been of more direct benefit to schoolmen in other cities than to those in the city for which it was made.

Thoughtful superintendents have come to see that, above everything else, the best insurance against the survey lightning is a survey conducted from within. Many a schoolman has asked himself the searching questions: "What would a survey reveal about my school system? On a factual basis, what do I really know about my own work?" Straightway he has begun himself to investigate. The various types of studies made in a modern survey are repeated on a smaller scale. Measurements are made and constructive attempts at remedial adjustments are begun.

The average administrator, however, is a busy man. He has little time, and by virtue of his training and position, little aptitude for close analytical studies or statistical investigation. Ordinarily he himself does just enough to realize the importance and value of the work, then delegates it to a specially trained person who works under his direction. In the smaller towns and cities this person may be only a clerk or bookkeeper who performs the mathematical and statistical labors involved. In the medium-sized cities a supervisor or assistant superintendent is often detailed to give part or all of his time to the work, and in the larger cities there are formal organizations of special research departments.

One of the best formulations of the functions of such research work is to be found in the report of the Committee on School Inquiry, New York City. Professor E. C. Elliott in his study of the administration of the New York Schools[1] made among others the following recommendation:

"Recommendation III

"That there be established as an integral part of the system of school control, a Bureau or Division of Investigation and Appraisal.

"This bureau or division should be in charge of a chief or superintendent, who is directly responsible to the Board of Education, and should be organized in such a manner as to enable it to serve as the central agency for the gathering and interpretation of statistical and other data with reference to the schools; and also for the carrying on of such investigations as are necessary for the rational development and expansion of the school system. It should bear the same general relation to the Department of Education as the existing Bureau of Municipal Investigation and Statistics bears to the Department of Finance.

"The following arguments may be indicated:

"(1) The school system of the city suffers from a lack of definite, detailed knowledge of its own working and its own cost. As has already been pointed out, the fundamental importance of the inspectorial form of control has been recognized only to a very limited extent. And even where its importance is recognized, officials charged with the responsibility for administrative or supervisory duty appraise their own performances. Investigation that is needed is not carried on at all.

[1] *Report on Educational Aspects of Public School System of the City of New York to the Committee on School Inquiry of the Board of Estimate and Apportionment,* Volume II, page 401.

"(2) It is evident that one of two things will result in the immediate future. Either the work indicated for this proposed bureau will be attempted by agencies outside of the school system or else there must be established, within the school system, as an integral part of its organized control, an agency properly equipped with trained investigators to set forth to the supervisory and administrative officials of the school system, and the people of the city, those essential facts absolutely necessary for the intelligent development of schools and of public sentiment. Of these alternatives, it would seem that the latter is to be greatly preferred. No outside agency could carry forward the work of inspection and of formulating impartial judgments of results, and of proposing new procedures without much friction and loss of energy.

"(3) The problems of public education in New York City are not conventional problems. Many of the more pressing ones are new in the social and educational world. They cannot be solved by preconceptions, or the showing of hands. In so far as possible, the situation and causes that have generated these problems must be weighed and analyzed before rational and permanent solutions can be found."

This recommendation was immediately acted upon. A Division of Reference and Research was formally organized and began its work at the opening of schools, September, 1913. The quotation above, therefore, is of historical interest. For while the New York Bureau by no means represents the beginning of systematic, scientific study of school problems by school authorities, it was undoubtedly the first to be definitely organized for that sole purpose.

The real beginning of the movement that has led to the establishment of bureaus of research cannot be clearly traced. In some cities there have been for many years committees, bureaus and special commissions for the more or less systematic study of buildings, of children, of teachers, of instruction, and of many other forms of school activities. As rapidly as the movement for measurement has developed, the investigations of these agencies have become more and more truly scientific. Even today the organization and function of a bureau of research are not clearly defined, and all sorts of studies are being carried on by all types of workers. In some cities, the department consists of little more than a high-sounding title bestowed as a compliment upon some existing school officer. In others, the director of the bureau is given the rank and salary of an assistant superintendent and is a real director of an extensive de-

partment. Between these two extremes every type of variation may be found.

It is not surprising, therefore, that the actual work of a bureau of research varies from city to city, being determined in the main by the tastes and training of the director, and somewhat by local needs. In some cities attention has been given almost wholly to a study of costs and to purely administrative problems; in other cities the energy of the department has been expended only on the measurement of educational products; in still others the deliberate attempt has been made to do some work in every field. In general, the tendency seems to be to regard the special function of the department as the devising, giving, scoring, tabulating and interpreting of standard tests and the prosecution of such other investigations only as may aid in the interpretation of the results secured.

A city department of research ordinarily consists of a director and clerical or stenographic assistants. At first, the tendency was to select any capable schoolman available, with little regard for his qualifications for the special work, but later appointees, either as directors or assistants, have been young men or women specially trained in statistical methods and in educational measurements. The director, almost without exception, is responsible directly to the superintendent and under his immediate control. The salaries paid range from $1,100 to $6,000, (median of 13 cases $2,700).

It has proved very difficult to obtain complete information with regard to the number of cities which have organized bureaus or which are carrying on organized work in measurement.[2] In spite of repeated questionnaires and persistent efforts, only a most tentative list can be given. However, formal organizations are found in the cities listed on the following page.

No attempt has been made to list cities in which research work is being carried on without formal organization, because of the impossibility, both of making the list at all complete and of distinguishing between a mere incidental use of tests once from idle curiosity and their persistent, intelligent use, year after year, for worthy ends. Seattle, for instance, does all the work and has all the benefit of a bureau of research, but has no formal organization. The assistant

[2]The various functions of a bureau of research are illustrated in other chapters, and the discussions will not be repeated here.

PARTIAL LIST OF CITY BUREAUS OF RESEARCH

City	Title of Bureau	When Organized	Name of Director
1. Baltimore, Md.	Bureau of Statistics and Research		Edwin Hebden
2. Boston, Mass.	Department of Educational Investigation and Measurement	1914	Frank W Ballou
3. Buffalo, N. Y.	Bureau of Research	1916	Wm. A. Mackey
4. Chicago, Ill.	Department of Standards and Statistics	1917	S. B. Allison
5. Cleveland, Ohio	Department of Reference and Research	1916	C. W. Sutton
6. Detroit, Michigan	Department of Educational Research	1914	S. A. Courtis
7. Hibbing, Minn.	Department of Educational Research	1915	J. W. Richardson
8. Kansas City, Mo.	Bureau of Research and Efficiency	1914	Geo. Melcher
9. Louisville, Ky.	Psychological Laboratory	1914	Henrietta V. Race
10. Los Angeles, Cal.	Division of Research	1917	Robert Lane
11. New York City	Bureau of Research and Reference	1913	E. A. Nifenecker
12. New Orleans, La.	Bureau of Educational Research	1912	(Discontinued)
13. Oakland, Cal.	Bureau of Reference and Research	1914	Virgil E. Dickson
14. Omaha, Neb.	Bureau Educational Research	1917	H. W. Anderson
15. Rochester, N. Y.	Efficiency Bureau	1912	J. P. O'Hern
16. St. Paul, Minn.	Bureau Research and Efficiency	1917	L. L. Everly
17. Schnectady, N. Y.	.....	1913	H. L. Davenport
18. Topeka, Ks.	Bureau Research and Efficiency	1916	Ira J. Bright

superintendent in charge is a director of educational research in all but name. There are a large number of such cities throughout the United States. At the other extreme is the superintendent or teacher whose curiosity is stimulated by some talk or article, and who gives a test once to a single class. The sales of standard tests have grown to very great proportions. Last year, of a single popular test, nearly 900,000 were used, and the annual sale of a few other tests run well over 100,000 copies each. Nor is the use of tests confined to this country. Shipments are made to all quarters of the world. It seems quite probable, therefore, that the number of bureaus of research are destined to be greatly increased in the immediate future.

CHAPTER VI

COOPERATIVE WORK FROM A UNIVERSITY CENTER

ERNEST J. ASHBAUGH
Director Educational Service, Extension Division, State University of Iowa

One of the most significant developments in the university in the last twenty years has been the growth of the idea of service—service not only to the comparatively few who were able to come to the campus, but service to the many who were not able to come. The idea was developed largely in state universities, where the funds for support came from direct taxation of all the people of the state and where the leaders came to realize that the service of the university should be extended to all who were taxed to support it. Only recently, however, has the university as a whole made an organized effort to educate all the people in its territory and to render service in the solution of the problems of its supporters off the campus. In fact, some of our state universities are doing almost nothing in these lines at the present time.

One of the very latest lines of service to be developed is that of cooperative educational research. The movement is so completely in its infancy that the method and the organization of the work have been but imperfectly worked out. However, two aims are rather definitely agreed upon by those engaged in this work, namely: (1) to make the university bureau a center for the direction of cooperative work with the school people of the state in the solution of the problems in which the latter are most interested; (2) to make the bureau an agency for the collection of rough data; for the tabulation, organization, and interpretation of these data; and for the distribution of the results of the study to the people contributing and to others, in order that significant facts may be known by the workers in the field.

A third aim might be added (though it must necessarily be subservient to the other two)—that of gathering masses of scientific

data which may be used in the further study of educational problems.

Nearly all problems connected with public-school work will lend themselves to cooperative research when once the university and the school people in the state come to a close understanding. The early efforts in the collection of data on school attendance, age-grade and age-progress studies, school health and causes of absence, the use of standard tests in arithmetic, reading, spelling, etc.; the figuring of unit costs of instruction and scoring of buildings and physical equipment—together represent but a small portion of the field that may be entered with profit.

The length of the period of compulsory schooling is increasing, but it does not suffice merely to attend longer. Much needs to be known concerning the extent to which communities and states are fulfilling their obligation to *all* their children—the obligation of giving them seven to twelve years of educational opportunity of a kind which will function in self-support and a wholesome social attitude. We need to know not only the extent to which this opportunity is offered, but also the extent to which this opportunity is being utilized. Cooperative research from a university center could obtain this knowledge.

We are in a time of critical questioning concerning the materials of education and the quality of the product turned out by our schools. The whole field of the school curriculum is open for research. Experimental methods certainly ought to be used in determining such things as time allotment, material to be used in each of the various grades and the arrangement of this material throughout the course. It is possible that method of presentation might also be included. This experimentation could be advantageously directed in cooperating schools by the university research bureau. Each problem would be specifically stated and the technique of procedure worked out at the university. The bureau would have the constructive criticism of the faculty in education and the preliminary experimentation would be carried on in the university experimental school. This would prepare the way for most effective work in the public schools which were cooperating with the bureau. As results were secured, each advanced step would thus be-

come immediately available to all the schools of the state. Many city superintendents are experimenting along some of these lines, but their experiments are not coordinated. Cities having bureaus of research are doing more extensive and critical work. But under the leadership of the university it would be possible to conduct such experiments under the varied conditions of small and large city, with the typical schoolroom conditions of each, and thus render a service to all communities in the state.

It is not the purpose of this chapter to list all the problems of cooperative research nor to give in detail the manner of solution. The purpose is rather to state the aims of a research bureau in a university center and to note very briefly some of the larger fields that await attack.

Development of the Iowa Bureau of Educational Service

Method—The fundamental aims noted above have been kept constantly in mind in the development of the Bureau of Educational Service in the Extension Division of the University of Iowa. The work was started in the fall of 1914, and the writer has been in charge of the work since its beginning. Some time was spent in traveling over the state and talking with the superintendents and thus learning at first-hand the most promising lines for initial endeavor.

The field had been partially prepared by courses in tests and measurement given in the College of Education during the preceding two years. Schoolmen attending the University had gained some knowledge of school surveys and some interest in cooperative work. The hearty cooperation of the faculty in education has been one of the greatest assets of the Bureau. In classroom, in institutes, and at teachers' associations they have furthered the work through frequent reference to the Bureau and its activities. The director of the Bureau holds rank in the College of Education and is a member of the instructional staff during the summer session. He thus comes in contact with the superintendents and principals who take summer work and assists in training them for further work in the field.

The annual conference on supervision, held at the University under the joint auspices of the Extension Division and the College of Education, has contributed much to the development of the work. It has brought together a group of school people (increasing in number from year to year) who are primarily interested in the problems of supervision. Such men as Judd, Strayer, Coffman, Ayres, Bagley and others have appeared upon the programs and promoted the interest in scientific education. The director of the Bureau has had a place at one general session each year and presented the results of research studies. At this conference, at the State Teachers' Association and at the sectional educational meetings each year, the director has met superintendents, principals and teachers in general sessions and round-table discussions. He has also met another group of men and women, the county superintendents. These persons, almost the only supervisors of the rural teachers, are becoming more and more interested in measurement of results, and their cooperation is being secured.

In a word, the secret of the development of the Iowa Bureau has been in the establishment of cordial cooperative relations with the various educational agencies of the state.

Fields—The first research problem attacked was that of attendance and the second,the measurement of school progress through the use of standard tests. In February, 1915, an arrangement was made with Mr. S. A. Courtis whereby the Bureau has exclusive control of his Series B tests in the state. In November of the same year a similar arrangement was made for the handling of the Kansas Silent Reading tests. These arrangements are still in force. Meanwhile, the Bureau has kept a small quantity of other tests on hand and stood ready to secure any others upon request. In addition to furnishing the superintendents of the state with the test material at cost, the Bureau has stood ready to give personal service in the giving of these tests. The writer has gone into a number of schools and demonstrated to the superintendents and principals the method of giving, scoring, and interpreting the results of these tests. This has resulted in the schoolmen becoming very much interested in the new field of tests and measurement. Many of these men and women have since studied at the University to increase their knowledge along these lines.

A third activity developed has been an information service by which the Bureau attempts to give, through its contact with the various departments of the University, technical information on any problem connected with the schools of the state. Questions on buildings, heating, lighting, ventilation, playgrounds, health service, course of study, census, finance, etc., have come from school boards, superintendents, teachers and patrons, and the Bureau has transmitted to the inquirers the best information available.

Another feature of the work of the Bureau has been local school surveys. On joint invitation of the school board and superintendents, a survey of any phase of a local school problem will be carefully made, the results analyzed and recommendations rendered. Care is taken here not to encroach upon the field of legitimate private enterprise. For example, service that belongs definitely to the field of an engineer, architect, or public accountant is not given.

Present Status of Work

As a result of the activity of the Bureau during the three and a half years of its existence, its status has been quite firmly established. At first there was a question in the minds of many, both in and out of the university, whether such a bureau was a legitimate part of the activity of a university and perhaps even more of an Extension Division. At present the doubt no longer exists in Iowa. The Bureau, through its service to the superintendents and school boards, has settled this question affirmatively.

It is understood at the University that it is the function of the Bureau of Educational Service to conduct researches in the field of education looking toward the promotion of efficiency in school work. All reasonable assistance will be given by various departments and colleges of the University to the furtherance of this work. This cooperation has been secured through a sincere effort to ask only reasonable assistance and to give full credit to the college, department or individual which rendered assistance to the Bureau.

The relation of the school people of the state to the Bureau has always been that of voluntary cooperation. No system of university credit for work done or other means of stimulating their cooperation has been offered. At all times there has been a definite under-

standing between the director of the Bureau and the superintendents that each superintendent should decide for himself whether he ought to cooperate in any proposed research. If he believed that the result of the research would be sufficiently valuable to his own school to justify the expenditure of the time and effort required, the Bureau would be very glad to receive his contribution. If he did not believe this would be the case,the most friendly relationship was maintained, and the invitation was repeated when another line of research work was undertaken. Thus, each superintendent in the state has come to look upon the Bureau as asking for only such assistance as shall contribute directly to the solution of his own problems.

Results Accomplished

The results accomplished fall rather definitely into three groups, according as they pertain to (1) state-wide surveys, (2) local surveys, and (3) general service.

State-wide Surveys—

(1) A state-wide survey of handwriting involving 110 cities and towns and rural pupils from fourteen counties was made in 1915 and the results issued in bulletin form.[1] The following six questions were asked and answered on the basis of the information secured by the survey:

1. How well do Iowa school children write?
2. Do children improve their quality of writing regularly as they progress through the grades?
3. Do children attending school in towns and cities write better than those attending the rural schools?
4. Do the children in the larger cities write better than those in towns or smaller cities?
5. How do children in this state compare with children in other states?
6. Is the quality of writing of the average eighth-grade child sufficient to satisfy the ordinary demands of every-day life outside of school?

The samples of writing were scored by the Ayres Handwriting Scale[2] and the conclusions were as follows:

[1]E. J. Ashbaugh, Handwriting of Iowa School Children, *Extension Bulletin, No. 15, State University of Iowa.*

[2]Leonard P. Ayres, *A Measuring Scale for Handwriting,* Russell Sage Foundation.

1. Iowa school children in the eight grades write approximately at Qualities 30, 35, 40, 45, 50, 52, 57, and 60, respectively, on this scale. Reference to the scale is necessary to understand the means of these values.

2. Yes. The improvement is quite uniform through the lower grades, but less rapid in the upper.

3 and 4. No. The differences between the quality of writing of children attending the rural school, small town and cities are negligible.

5. On the average, Iowa children are writing as well as children of like grade elsewhere in the United States.

6. Apparently the quality of writing of the majority of eighth-grade children will satisfy the ordinary demands of daily life, since 75 percent of these children write a better quality than is required by the New York Municipal Civil Service Commission.

(2) A similar survey of achievement in the fundamentals of arithmetic as measured by the Courtis Series B tests was made in 1916 and the results distributed in a bulletin.[3] The following four questions were proposed and answered in this bulletin:

1. How skillful are Iowa children in performing the four fundamental operations in arithmetic?

2. How does the skill of Iowa school children compare with that of children of like grades in other states?

3. How does the skill of children in small towns compare with that of children in larger towns and cities?

4. What use can be made of Standard Tests?

1. The median speed and accuracy of children were ascertained in each of the grades. Mr. Courtis' standard as well as the scores of each of the cities contributing data was furnished for purposes of comparison. The evidence indicated that more speed was needed in the upper grades and greater accuracy in all.

2. Iowa school children were shown to excel in most grades and operations when their scores were compared with the available records of sister states.

3. While, in general, the scores of pupils in smaller towns are lower than those of pupils in larger places, the records of some small towns showed clearly that size of place is not a determining factor.

On the basis of these showings many schools in the state have modified their courses in order to give greater attention to drill work in fundamentals.

---

[3]E. J. Ashbaugh, Arithmetical Skill of Iowa School Children, *Extension Bulletin, No 24, State University of Iowa.*

(3) Less elaborate studies have been made of achievement in reading and spelling, attendance, causes of absence and teachers' marks. In reading and spelling, the general state situation seems to be average or slightly above when compared with other states. Individual cities have discovered weaknesses through these surveys and modified their practice accordingly. The attendance survey showed an amount of absence almost incredible to superintendents and teachers—20 percent of the children out more than 10 percent of the time; 10 percent of the children out more than 20 percent of the time; 6.6 percent of the children out more than 30 percent of the time—not a satisfactory situation. High 'percentages of attendance' have been secured through the method of dropping a child's name from the roll after a period of a day and a half to three days' absence. This, however, does not make for increased school attendance by the child.

In all these studies the random selection of cities and the large number of pupils involved makes it probable that the results are typical of the state. Definite readjustments are known to have been made in a number of school systems on the basis of the facts revealed.

Local Surveys—

(1) In response to direct invitations by superintendents and school boards, more or less complete surveys of four Iowa school systems have been made. Written reports have been rendered to the school boards in each case.

(2) Building surveys involving the question of utilization of present buildings and the best solution for caring for the increased school population have been conducted in three cities.

Local surveys are possible only by the closest cooperation with teachers, superintendents and school boards. Meetings with teachers are arranged where the results of the surveys of instruction are given and the interpretation is carefully explained. These meetings enable the teachers to apply the survey results so as to increase classroom efficiency. Meetings with the school board are also arranged so that information may be given them that will help them to fulfill their functions to the greater good of the schools.

General Service—

(1) The various standard tests have been furnished at cost to the schools of the state. This has kept the Bureau in touch with all schools doing measuring work and fostered the cooperative relationship.

(2) An informational service has been established which endeavors to be a source of help in the solution of school problems. This service is free and open to anyone within the state.

The response of the school superintendents of the state to any projected cooperative work, the large number of calls for surveys of various kinds, and the utilization of the obtained results, furnish the best evidence of the value of the work that has been accomplished. The fact that calls are still coming in for studies issued more than two years ago indicates that the work of the Bureau is considered as having more than temporary value. During the three years and a half cooperative relationship has been established with more than 100 different cities and towns of the state. At the present time I know that when an invitation is extended for cooperative work, a truly representative number of cities and towns can be relied upon to furnish the desired information.

Difficulties

A fundamental difficulty, and one that often tends to invalidate results, is that a large number of teachers, principals and superintendents have not had an opportunity for training in scientific research work. With the very best intention to cooperate, directions are frequently misunderstood or the importance of their being followed in an absolute manner is not appreciated and thus a variable factor enters into the work. This difficulty can be overcome only by extreme care and preliminary experimentation in the formulation of directions.

Another problem, perhaps even greater than the securing of reliable data, is presented by the question: How shall the results of researches be reported in such a way that the greatest possible good may come to schools? The purpose of the research is the modification of schoolroom practice. In most cases this will be accomplished only when the results reach the teacher in a form which

is clear, definite and usable. It must be remembered that the movement in scientific education in recent years is demanding a new type of supervisor—one trained in the handling and interpretation of scientific data. But until the universities, colleges and normal schools are able to give us this trained corps, and a representative is found in every school, it will be necessary to depend largely upon the classroom teacher for the application of the report to practice. Hence, special effort should be made, first, to present the report in such form that the ordinary teacher may clearly see its meaning and be inspired to utilize the results in her own classroom, and second, to distribute the report in such a manner that it shall come directly to the attention of the teacher herself.

The difficulty of securing adequate assistance in caring for the work presents itself to nearly every bureau. To get reliable data in many studies it is often more necessary to score the papers than to give the tests. This requires much time and clerical assistants with a specific kind of training. Where the director of the bureau is on the regular teaching staff of the institution, he may personally solve this difficulty by the use of research data as laboratory material with his classes. This enables him to train his help and give his students very valuable experience with this kind of work at the same time. A counter-difficulty arises with this plan in that the director is not free for extended absences over the state as calls may require.

State surveys that involve the collection of material in many schools and the compilation and interpretation of results at the central bureau present only the difficulties of securing the needed cooperation in the field and the necessary assistance at the bureau. These can be met. But local surveys present a well-nigh insuperable difficulty in the element of time. Even with a large staff, extensive local surveys could be made in only a few cities in a year. Fortunately, many features of a local survey can be done as well by the local superintendent, if he be a trained man, as by a representative of the bureau. Hence, a partial solution to this difficulty lies in the training of superintendents.

Suggestions

On the basis of three years and a half experience with the Bureau at the University of Iowa, of frequent conversations with those in charge of similar bureaus elsewhere, I offer the following suggestions to those who contemplate the creation of a similar bureau:

1. The bureau should be so financed as to make available sufficient funds to care for the immediate tabulation of the results of any study conducted by it.

2. A plan should be devised by which the director of the bureau and persons interested in its cooperative activity may meet for consultation one or more times during the year. These meetings might be arranged in connection with the State Teachers' Association or other educational gathering at which a large group of school superintendents would be present.

3. The director should be a member of the teaching staff of the university during its summer session so that through the classroom he may increase the interest in the work of the bureau and assist in the training of teachers, principals and superintendents for educational investigation.

4. Results of researches should be presented to the public, in a clear and forcible manner, attractive to the lay reader and in a form usable by the average teacher. Specific problems should be set up to which specific answers are given. These problems and their answers at least should be intelligible to all readers, even though some readers may not be able to follow clearly all the processes by which the results are obtained.

5. With the issuance of the studies of the bureau, arrangements should be made for a publicity campaign that will focus the attention of the school people upon these results.

Historical Note

Pioneer efforts along the line of cooperative investigation in educational measurements were made in 1910 by S. A. Courtis in arithmetic. One of the outgrowths of his activities was the taking over of the research work within states by state universities. The first university to recognize the possibilities of service through the

formation of a bureau of cooperative research was the University of Oklahoma, in 1913. At this time, however, cooperative work was also under way in the University of Indiana, and the formal organization of the Bureau of Cooperative Research took place the following year. Similar bureaus at the University of Iowa and at the State Normal School, Emporia, Kansas, were organized a little later. A partial list of university bureaus of research (with sufficiently formal organizations to have stationery of their own!) is as follows:

University	Name of Bureau	When Organized	Name of Director
Univ. of Arkansas, Fayette, Ark.	Bureau of Educational Tests and Measurements	1917	J. R. Jewell
Univ. of Indiana, Bloomington, Ind.	Bureau of Cooperative Research	1914	(Position unfilled at present)
Univ. of Iowa, Iowa City, Iowa	Educational Service, Extension Division	1914	E. J. Ashbaugh
Univ. of Kansas, Lawrence, Kan.	Bureau of Educational Measurements	1916	F. J. Kelly
Univ. of Minnesota, Minneapolis, Minn.	Bureau of Cooperative Research	1915	M. E. Haggerty
Univ. of Nebraska, Lincoln, Neb.	Bureau of Educational Measurements	1914	Charles Fordyce
Univ. of Oklahoma, Norman, Okla.	Bureau of Measurements and Efficiency	1913	W. W. Phelan
Univ. of South Dakota, Vermillion, S. Dak.	Bureau of Educational Research	1915	W. Franklin Jones

It must not be supposed, however, that the universities that appear in the list above are the only ones actively supporting the movement. From Harvard and Columbia Universities in the east, to Leland Stanford Junior University in the west, from the University of Wisconsin in the north to the University of Texas in the south, similar work is being done by schools and departments of education. If a distinction is to be made at all, it is that the universities that have formal bureaus usually act as distributing centers for testing material, but even this distinction does not always hold. The influence of university men in education has naturally been one of the main factors that have led to the growth of the movement.

The university, moreover, has by no means been the only factor. Normal schools and teachers' institutes have done their share. At present bureaus of research are to be found in at least two normal schools:

Normal School	Name of Bureau	When Organized	Name of Director
State Normal School, Emporia, Ks.	Bureau of Educational Standards and Measurements	1914	Walter S. Monroe
Northern Normal and Industrial School, Aberdeen, So. Dak.	Bureau of Educational Research	1917	Willis E. Johnson

Still another agency for the development of educational research throughout a state has been the state department of education. In at least three of these attention to research work is the special duty of a particular member of the staff.

State	Name of Bureau	When Organized	Name of Director
New York	State Dept. of Education		Wm. A. Averill
Wisconsin	State Dept. of Education	1915	W. W. Theisen
Georgia	State Dept. of Georgia		M. L. Duggan

Probably this is the proper place, also, to comment on the aid rendered by the U. S. Bureau of Education. Bulletins and reports of committees on standard tests and scales have been printed; surveys have been conducted, and in many ways the Bureau at Washington has done what it could to further the cause of measurement.

Finally, tribute must be paid to the influence of the great foundations. The Russell Sage Foundation, through its Division of Education, under the direction of Leonard P. Ayres, was the pioneer in this field and has been one of the major influences responsible for the rapid development of the movement for measurement. The Ayres Scales in writing and spelling are widely used, while the investigations of the Division in the field of cost-accounting and child-accounting have had even greater influence. The survey work carried on by the Division has been a third type of activity that has had a great effect in directing the minds of school men to the possibilities of measurement. The contributions of the foundation to this field are very great.

Of recent years, more and more attention has been given by other foundations to the field of educational research. The Cleve-

land Foundation, although of local origin and interest, has, through the Cleveland Survey, made its contribution to the educational progress of the country. The Carnegie Foundation has recently entered the field.

The General Education Board, also, has given many evidences of its interest in the scientific study of educational problems. It has supported the New Hampshire Bureau of Educational Research, subsidized research work at the University of Chicago and in other places, conducted surveys in the Maryland and Gary schools, and is now carrying on an experimental school.

It seems clear, therefore, that in the future the development of educational measurement should be even more rapid than in the past. At one extreme are great foundations willing to expend large sums of money for any educational investigation that promises to yield results of permanent value. At the other extreme is a vast army of educational workers actively engaged in teaching children and anxious to make use of every tool, device, or method, that will help them to do better work. Between the two are the universities, colleges and normal schools, training men both to carry on research work successfully and to apply the results of experimental studies to the practical problems of the schoolroom.

CHAPTER VII

EXISTING TESTS AND STANDARDS

WALTER S. MONROE
Director, Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas

During the past decade and especially during the past five years, the number of tests available for measuring the abilities of children in school subjects has grown very rapidly. Those described in this chapter are as follows:

STANDARDIZED TESTS FOR USE IN THE ELEMENTARY SCHOOL

Arithmetic	17	Language	17
(Fundamental Operations 11)		Music	1
(Arithmetic Reasoning 6)		Silent Reading	13
Drawing	1	Oral Reading	4
Geography	6	Spelling	11
Handwriting	10	Total	84
History	4		

STANDARDIZED TESTS FOR USE IN THE HIGH SCHOOL

Algebra	7	History	1
Drawing	1	Physical Training	1
Foreign Languages	11	Physics	1
Geometry	3	Total	25

The superintendent or teacher who wishes to measure the results of instruction, faces the problem of making a wise choice from this material. This account of existing tests and standards has been prepared to assist in making this choice. Because of the limitations of space, only brief descriptions are possible. In the chapter prepared by Miss Bryner the reader will find references to the accounts of the derivation of all the tests and of their use, whenever they are available. Most of the tests are also described more completely in the writer's book on *Educational Tests and Measurements*, Houghton Mifflin & Company, 1917.

The standards given in several cases are tentative only and all are, of course, subject to revision. Unless standard scores are based upon a large number of cases, they should be used with caution in making comparisons. In all cases it should be remembered that any standard which is simply a statement of the concensus of present practice is open to the criticism that we must not assume that what is, is what should be. For example, it may be that a high degree of efficiency requires a much higher standard in the rate of silent reading than the average rate at which children now read in the several grades.

It should also be remembered that the scores upon which these standards are based were obtained by following certain definite directions in giving the several tests. Even slight variations in procedure often materially affect pupils' scores. Hence, when comparison with standard scores is the object of the testing work, one should follow the standard, or specified, directions which in most cases accompany the tests. In any case, in making comparisons it is unwise to attach great importance to small differences in scores.[1]

An effort has been made to make this list as complete as possible, but doubtless some tests have been overlooked, and it is certain that within a short time new tests will be announced. In fact, several were found that were in the process of derivation. Certain special tests which have been devised for laboratory research are intentionally omitted because they are not available for distribution and in general are not suitable for the type of testing the superintendent or teacher will do. Practice tests or other exercises which are primarily teaching devices have also been omitted.

The tests may be obtained from the addresses given When no address is given, the tests are not available for general use. Even in a few cases where the address is given, the tests are not available. On account of recent fluctuations in the cost of printing, prices are not stable. Accordingly, no prices are given. However, in only a few cases are the tests published on a commercial basis and for this reason one may be reasonably certain that the tests may be obtained at approximately the cost of printing.

[1]S. A. Courtis, *Third, Fourth and Fifth Annual Accountings, 1913-1916.* (Department of Cooperative Research, Detroit). Read especially the warning given on page 53.

Standardized Tests for Use in the Elementary School

I. Arithmetic, Fundamental Operations

1. Bobbitt's Arithmetic Tests. These tests were used in the survey of the public schools of San Antonio, Texas. They consist of nine tests, five for the operations with integers and four for the operations with common fractions. Each test is limited to one operation, but contains examples of several different types. However, they are arranged in groups which are equal in difficulty. (Ref. 546.)*

2. Boston Tests. Addition of Fractions. These are a series of six tests devised by A. W. Kallom and are significant for the illustration they furnish of tests based upon a scientific analysis of the abilities they measure. This analysis revealed fourteen types of examples in the addition of two fractions, but by making certain combinations the number of tests needed to measure this group of abilities was reduced to six. A similar series has been devised for subtraction of fractions, and it is planned to extend the work to multiplication and division. (Ref. 78.)

Standards: Boston Medians: Addition of Fractions

Grade	Pupils Tested	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6	
		Speed Medians	Accuracy Medians	Speed Medians	Accuracy Medians	Speed Medians	Accuracy Medians	Speed Medians	Accuracy Medians	Speed Medians	Accuracy Medians	Speed Medians	Accuracy Medians
VI	1205	10 7	79.6	7.7	65.6	5.5	41.9	4 0	69.5	4.6	51 0	4 4	48.6
VII	1243	16.5	86.6	10.1	72.9	7.3	46.1	5.3	69.2	6.3	54.9	5.7	48.1
VIII	1130	20.7	88.2	11.6	74.4	8.4	47.4	6.0	67.8	6.9	52 4	6.4	46.5

3. Cleveland Survey Tests. These were designed for use in the survey of the Cleveland Public Schools. They have been revised slightly and used in the surveys at Grand Rapids and St. Louis. The series consists of fifteen tests, including four in addition, two in subtraction, three in multiplication, four in division, and two in addition and subtraction of common fractions. The total working time is 22 minutes and the administration of the tests is simple. They furnish a more detailed analysis than can be secured by means of the Courtis Standard Research Tests, Series B. Address Charles H. Judd, School of Education, University of Chicago, Chicago, Illinois, or S. A. Courtis, 82 Eliot St., Detroit. (Ref. 403.)

*Reference numbers in this chapter refer to the numbers in the bibliography, Chapter XIII.

AVERAGES OF MEDIAN SCORES IN 15 ARITHMETIC TESTS FOR GRADES 3-8. CLEVELAND AND GRAND RAPIDS. NUMBER OF EXAMPLES RIGHT

Test	Grade					
	III	IV	V	VI	VII	VIII
A	13.4	17.1	21.9	24.9	27.0	28.9
B	8.9	12.8	16.6	19.5	21.1	25.8
C	6.5	11.7	14.8	16.8	18.2	19.9
D	6.3	11.4	15.0	17.7	20.3	22.8
E	4.3	5.0	5.9	6.7	7.4	8.0
F	2.0	4.5	6.6	7.7	9.1	10.6
G	2.0	3.6	5.1	5.5	6.0	6.7
H	...	...	5.6	6.0	7.7	8.6
I	0.6	1.0	1.7	3.1	4.0	4.7
J	1.9	3.0	3.9	4.4	5.1	6.1
K	...	4.0	5.6	7.0	9.4	11.4
L	...	1.7	2.7	3.2	3.8	4.4
M	1.4	2.4	3.4	4.1	4.7	5.4
N	...	0.8	1.1	1.6	1.9	2.4
O	...	...	...	3.3	4.3	5.2

4. Courtis Standard Tests, Series A. This series includes eight tests, one for the combinations (0-9) in each of the operations; copying figures, speed reasoning, fundamental operations, and reasoning. The series was devised in 1909 and was used extensively during the following years. However, the author has discontinued its publication in favor of Series B, devised in 1913. (Ref. 91.)

STANDARD MEDIAN SCORES, COURTIS' STANDARD RESEARCH TESTS, SERIES B.

GRADE	Addition		Subtraction		Multiplication		Division	
	Speed	Acc.	Speed	Acc.	Speed	Acc.	Speed	Acc.
IV—General	7.4	64	7.4	80	6.2	67	4.6	57
Courtis	6	100	7	100	6	100	4	100
Boston	8	70	7	80	6	60	4	60
V—General	8.6	70	9.0	83	7.5	75	6.1	77
Courtis	8	100	9	100	8	100	6	100
Boston	9	70	9	80	7	70	6	70
VI—General	9.8	73	10.3	85	9.1	78	8.2	8
Courtis	10	100	11	100	9	100	8	100
Boston	10	70	10	90	9	80	8	80
VII—General	10.9	75	11.6	86	10.2	80	9.6	90
Courtis	11	100	12	100	10	100	10	100
Boston	11	80	11	90	10	80	10	90
VIII—General	11.6	76	12.9	87	11.5	81	10.7	91
Courtis	12	100	13	100	11	100	11	100
Boston	12	80	12	90	11	80	11	90

Speed is the number of examples done in the time allowed.

Accuracy is the percent of examples correct.

"General" medians were determined by Courtis on the basis of the 1916 tabulations and summaries of tabulations of other years. Courtis, S. A. *Third, Fourth, and Fifth Annual Accountings, 1913-16.* (Department of Cooperative Research, Detroit).

The Boston standards were established after using the tests for three years. Ballou, F. W., Arithmetic, the Courtis Standard Tests in Boston, 1912-15. (*Bulletin No. 10 of the Department of Educational Investigation and Measurement*)

5. Courtis Standard Research Tests, Series B. This series of tests consists of one test in each of the four fundamental operations. The tests measure the speed and accuracy with which the pupil can perform these operations with one type of example. The administration is very simple. The total time required to give them is 26 minutes. They have been used extensively since their first publication in 1914. The measures have been proved reliable in 75 to 90 percent of the cases. Address S. A. Courtis, 82 Eliot St., Detroit, Michigan. (Refs. 77, 97.)

6. Guhin's Number Tests. These tests include 88 combinations for both addition and multiplication. The standards are given in terms of the number of seconds required to complete the test. Address Hubb City School Supply Company, Aberdeen, South Dakota.

Standards.	3rd grade, 150	6th grade, 120
	4th grade, 140	7th grade, 110
	5th grade, 130	8th grade, 100

7. Monroe's Diagnostic Tests. This series covers the four fundamental operations in integers, common fractions, and decimal fractions. It is thought that they will furnish a reasonably complete diagnosis of the abilities of pupils to perform the operations of arithmetic. Although the series consists of 21 tests, they have been so arranged that the total time required for giving them is only 35½ minutes. Address Bureau of Educational Measurements and Standards, Emporia, Kansas.

8. National Business Ability Tests. The tests for addition, subtraction and multiplication are abbreviated forms of the corresponding tests of the Courtis Standard Research Tests, Series B. In addition to these three, there are tests in multiplication of common fractions and in percentage. The standards are stated in terms of the number of minutes allowed for completing the respective tests. Address Sherwin Cody, Managing Director, 189 W. Madison St., Chicago, Illinois.

9. Stone's Arithmetic Test for the Fundamental Operations. This test is of historical interest because it was used by Courtis

in the experimentation which resulted in the derivation of Series A. It was designed as a general test and has not been standardized except for the sixth grade. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 138.)

10. Thompson's Standardized Tests. This is an elaborate series of tests upon the operations of arithmetic. The feature of these tests is a mechanical device for scoring the papers. Address T. E. Thompson, Monrovia, California.

11. Woody's Arithmetic Scales. These consist of two series of four tests, one for each of the fundamental operations. They differ from such tests as the Courtis Standard Research Tests, Series B, in that the examples in each scale have been carefully graded and arranged in order of difficulty. In content they include integers, decimal fractions, common fractions and denominate numbers.[2] Series A and Series B are similar, except that Series A is more finely divided. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 148.)

TENTATIVE STANDARDS OF ACHIEVEMENT FOR WOODY TESTS, SERIES A

Grade	Addition	Subtraction	Multiplication	Division
II	3.12	1.44	...	...
III	4.99	2.96	1.89	2 54
IV	6.11	4.22	4.05	3 21
V	6.99	5.47	5.52	4 94
VI	7.95	6.46	6.72	5.87
VII	8.65	7.31	7.26	6.59
VIII	9.01	7.64	7.93	7.16

TENTATIVE STANDARDS OF ACHIEVEMENT FOR WOODY TESTS, SERIES B

Grade	Addition	Subtraction	Multiplication	Division
II	4.5	3	...	...
III	9	6	3.5	3
IV	11	8	7	5
V	14	10	11	7
VI	16	12	15	10
VII	18	13	17	13
VIII	18.5	14 5	18	14

[2]Recently there has appeared a modification of the Woody tests known as the Woody-McCall Mixed Fundamentals, Series B, I and II. These tests are more difficult than the original Woody tests. Each sheet has on it problems in all four of the fundamental operations, so that the pupil must choose the right operation for each problem. In exploring for tests best fitted for selecting gifted children in the 5th and 6th grades I have found the Woody-McCall Mixed Fundamentals distinctly better than the Woody tests from which they were derived.—G. M. W.

The standards are expressed in terms of the degree of difficulty of the examples that are done correctly by just 50 percent of the pupils.

II. Arithmetic: Reasoning

1. Bonser's Reasoning Tests. These consist of two lists of ten problems each. The problems have been chosen so that the two tests are equal in difficulty. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 75.)

2. Buckingham's Reasoning Tests. These tests were devised for use in the survey of the Gary and Prevocational Schools of New York City. The problems of the tests were carefully evaluated and arranged so that the two lists are equally difficult, but were not scientifically selected. (Refs. 82, 466.)

3. Courtis' Reasoning Tests. Tests 7 and 8 of the Courtis Standard Research Tests, Series A (*q. v.*) are reasoning tests.

4. Rice's Reasoning Tests. These were given by Rice in 1902 and are of historical interest. They consist of a series of tests, one for each of the grades from fourth to eighth, inclusive. The problems were selected as suitable for the pupils of the respective grades. (Ref. 129.)

5. Starch's Arithmetical Scale A. This test consists of a series of arithmetical problems which are arranged in order of increasing difficulty. Address Daniel Starch, University of Wisconsin, Madison, Wisconsin. (Ref. 134.)

The following are standard scores for the ends of the respective years, as derived from 2515 pupils in 18 schools:

Grade	III	IV	V	VI	VII	VIII
Score	4.5	6.2	7.8	9.4	11.0	12.6

6. Stone's Reasoning Test. This is a single test designed to be given to Grades IV to VIII, inclusive. The problems have been carefully evaluated. The test was used in the survey of the public schools of Butte, Montana, and Salt Lake City, Utah. Address Bureau of Publications, Teachers College, Columbia University, New York City.

STANDARDS. MEDIAN SCORES FOR THE STONE REASONING TEST

Grade	Butte	Bridgeport	Salt Lake City	18,495 Wisconsin Pupils
V	2.7	6.1	4.3	2.4
VI	4.4	5.2	6.9	3.9
VII	6.3	6.8	9.1	5.4
VIII	8.2	4.5	11.0	6.9

Recently Stone has issued the following standards:

"That 80 percent or more of 5th-grade pupils reach or exceed a score of 5.5, with at least 75 percent accuracy; that 80 percent or more of 6th-grade pupils reach or exceed a score of 6.5, with at least 80 percent accuracy; that 80 percent or more of 7th-grade pupils reach or exceed a score of 7.5, with at least 85 percent accuracy; that 80 percent or more of 8th-grade pupils reach or exceed a score of 8.75, with at least 90 percent accuracy." (Ref. 139.)

III. Drawing

1. Thorndike's Drawing Scale. This scale, devised by E. L. Thorndike, in 1913, consists of a series of drawings arranged in order of merit as determined by competent judges. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 159.)

IV. Geography

1. The Boston Tests. The two tests of this series—one on the United States and the other on Europe—consist of well-chosen questions. The relative difficulty of the questions was determined upon the basis of the percent of correct answers. The tests were devised in an effort to determine: (1) the character of achievement in geography and (2) the possibility of scientific measurement of educational results in geography. This significant comment is made: "The results show how inadequate the customary examination or test in geography is to measure ability in geography." (Ref. 216.)

2. Buckingham's Geography Test. This test was devised for use in the survey of the Gary and Prevocational Schools of New York City. It consist of two sets of 20 questions which were evaluated upon the basis of the percent of correct responses. (Ref. 466.)

3. Hahn-Lackey Geography Scale. This scale consists of several hundred geographical questions which were found to be com-

mon to six modern texts and which satisfied certain other criteria. These questions have been classified according to difficulty. In appearance the scale is very much like the Ayres' Spelling Scale and is to be used the same way. Address H. H. Hahn, Wayne State Normal School, Wayne, Nebraska.

4. *Starch's Geography Tests, Series A.* The common elements of five geography texts have been arranged in five parallel tests. The exercises of the tests are in the form of mutilated sentences. Address Daniel Starch, University of Wisconsin, Madison, Wisconsin.

5. *Thompson's Standardized Tests in Geography.* These consist of a test each for North and South America. They deal entirely with place geography. An important feature is a mechanical device for scoring the papers. Address T. E. Thompson, Monrovia, California.

6. *Witham's Standard Geography Tests.* These are a series of tests arranged to test quickly and easily pupils' knowledge of certain geographical facts. The facts for the tests on the world are grouped under these heads: (1) geographical divisions, (2) form and motion of the earth, (3) the hemispheres, (4) land and water forms, (5) homes of the races, (6) industries, and (7) largest cities. Address E. C. Witham, Southington, Conn. (Ref. 221.)

V. Handwriting

The scales described below are used to measure the quality of handwriting. The speed of handwriting is measured by having suitable material written under specified conditions for a definite number of minutes. In order that a measurement of speed may be most significant, it must be made when the quality of the pupil's handwriting is approximately standard.

"Pupils should be asked to write a suitable selection which they have memorized. To guard against lapses of memory, the pupils should be asked to repeat in concert the selection to be used. If convenient, it is well to provide each pupil with a printed or typewritten copy of the selection. When this cannot be done, the selection may be written on the blackboard where all can see it. The selection should contain no words which the pupils cannot spell

readily. It is well to have them practice writing the more difficult words before the test is begun. Do not use material which the pupils must compose as they write, for this would be worthless in testing. The rate of writing unfamiliar material from a printed copy will vary with the pupils' rate of reading and so will not give a true measure of speed. Dictated material should be used only when the teacher wishes to control the speed, not when speed is to be measured.

"Different investigators have required pupils to write different material. Several have used the first line or the first stanza of the poem, 'Mary had a little lamb.' 'Sing a Song of Sixpence' has been used. Other sentences which have furnished copy are 'Jolly kings bring gifts while happy maids dance.' 'A quick brown fox jumps over the lazy dog.'[3] 'Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small card. John vanished behind the bushes and the carriage moved along down the driveway.'[4] In the Cleveland Survey the first three sentences of Lincoln's Gettysburg Address were written, and Ayres has used the same selection in the 'Gettysburg Edition' of his scale. In several surveys the pupils were allowed to write any familiar stanza of a poem. The chief principles to bear in mind in selecting materials are: first, to use material in the lower grades which will not furnish difficulties in spelling and remembering; and second, to use material which will be uniform in all classes which are to be compared."[5]

1. Ayres' Scale for Measuring the Handwriting of School Children. This is known as the "Three Slant Edition," or more simply as the Ayres Scale. It consists of three types of specimens of the handwriting of school children—vertical, semi-slant and full slant—arranged in order of legibility as determined experimentally. The values 20, 30, 40, up to 90, have been assigned to the specimens. This scale has been used very widely. Address Russell Sage Foundation, New York City. (Ref. 227.)

[3]This sentence was used in securing specimens for the Freeman Scale. It contains all the letters of the alphabet.

[4]These sentences were used in securing the specimens for the Thorndike Scale.

[5]Walter S. Monroe, *Educational Tests and Measurements*, p. 146. Houghton Mifflin Company, 1917.

2. Ayres' Scale for Measuring the Quality of Handwriting of Adults. This scale is similar to that described above, except that specimens of the handwriting of adults are used instead of the handwriting of school children. Address Russell Sage Foundation, New York City. (Ref. 229.)

3. Ayres' "Gettysburg Edition." This scale differs from the other two in certain important characteristics. It consists of specimens of school children's handwriting on ruled paper and there is only one specimen for each division of the scale instead of three representing different degrees of slant. The copy, the first three sentences of Lincoln's *Gettysburg Address,* is the same for all specimens. The scale has printed on it standards for both speed and quality and complete directions for its use. Ayres asserts that the purpose of the new features is "to increase the reliability of measurements of handwriting." Address Russell Sage Foundation, New York City.

4. Breed and Downs' Scale. This scale was constructed in making a survey of the handwriting in the public schools of Highland Park, Michigan. The specimens were scored by means of the Thorndike Scale and then certain ones selected for a five-step scale for each of the following grades, 3d A, 3d B, 4th A, 5th A and 6th A. Thus, it differs from other scales in having a special scale for each of the grades named. (Ref. 234.)

5. Courtis' Standard Research Tests, Handwriting, Series W. Test I, Handwriting, is an untimed "maximum performance" test, designed to secure samples of the children's best writing after practice. Test II, Filing Test, is a "free-choice" copying test, designed to secure samples of the children's writing under working conditions. The test consists of the names and addresses of ten business firms, to be copied in alphabetical order. In both tests the quality of the writing is to be measured with the Ayres scale. The difference in quality between the two samples reveals any "lack of transfer" from the work of the writing class to ordinary writing. The material in the Filing Test has been so chosen as to afford excellent material for an analysis of the defects in the writing of a particular child. Address S. A. Courtis, 82 Eliot Street, Detroit, Michigan.

6. Freeman's Handwriting Scale. This scale differs from the others in that there is one scale for each of the following characteristics of handwriting: (1) uniformity of slant, (2) uniformity of alignment, (3) quality of line, (4) letter formation, and (5) spacing. Only three degrees of each characteristic 1, 3 and 5, are included in the scale, although the intermediate values, 2 and 4, may be used. This scale is designed for diagnosis rather than general measurement. Address Houghton Mifflin Company. (Ref. 240.)

7. Gray's Score Card. This score card is of the same general character as those which are used in judging grain and livestock. It is based upon a determination of the important characteristics of handwriting. Its function is similar to that of the Freeman scale, *i. e.*, to furnish a diagnosis rather than a general measurement. Address C. T. Gray, University of Texas, Austin, Texas. (Ref. 245.)

8. Johnson and Stone's Scale. This scale is similar in general plan to the Ayres and Thorndike Scales, but based on several factors, including movement and a detailed analysis of legibility. Each specimen of the scale is accompanied by a legend which states its defects and merits in terms of the analysis appended, which includes seven factors—letter formation, uniformity of slant, uniformity of alignment, spacing, quality of line, size, and degree of slant. (Ref. 247.)

9. Thorndike's Scale. This scale was constructed on the basis of three characteristics—beauty, legibility, and general merit. The degree of these characteristics represented in the specimens of the scale was determined by the concensus of opinion of competent judges. The numerical values of the specimens of the Thorndike Scale range from 4 to 18, and one or more specimens are given for each degree of quality. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 263.)

10. Zaner and Blossom Handwriting Scales. These are a series of scales for the several grades, designed to be used with a particular system of handwriting. Address Zaner and Blossom Co., Columbus, Ohio.

STANDARD MEDIAN SCORES. SPEED OF HANDWRITING

Source	Grades							Approximate number of specimens scored
	II	III	IV	V	VI	VII	VIII	
Cleveland[6]	39.2	...	...	60	70	76	80	25,387
Iowa Schools[7]	39.2	49.2	61.9	65.5	72.6	75	76.5	28,000
Starch's Standard[8]	31	38	47	57	65	75	76.5	 4,740
Kansas Medians[9]	32	35	51	61	67	71	73	6,000
Fifty-six cities[10]	30.6	43.8	51.2	59.1	62.8	67.9	73	34,000
Freeman's Standards	36	48	56	65	72	80	90	

STANDARD MEDIAN SCORES: QUALITY OF HANDWRITING

Source	Grades							Scale used	Approximate number of specimens scored
	II	III	IV	V	VI	VII	VIII		
Cleveland	....	....	....	45	48	50	55	Ayres	25,287
Iowa Schools	35.7	39.8	44.5	49.1	52.3	57	55	Ayres	28,000
Starch's Standard	27	33	37	43	47	53	57	Ayres	4,740
Kansas Medians	44	47	50	55	59	64	70	Ayres	6,000
Fifty-six cities	39.7	42	45.8	50.5	54.5	58.9	62.8	Ayres	34,000
Freeman's Standards (Ayres Scale)	44	47	50	55	59	64	70	Ayres	
Salt Lake City[11]	....	9.2	10.7	11.1	11.3	12.2	12.8	Thorndike	2,500
Butte, Montana[12]	8.2	8	8.8	8.9	11.6	11.2	12.1	Thorndike	1,400
Southington, Conn.[13]	....					10		Thorndike	1,200
Connersville, Ind.[14]	....	10.3	10	10.3	11.7	11.7	11	Thorndike	
Freeman's Standards (Thorndike's scale)	9.36	9.75	10.13	10.76	11.34	11.89	12.66		

VI. History

1. Buckingham's Tests. These tests were used in the survey of the Gary and Prevocational Schools of New York City. They consist of two sets of questions which have been evaluated on the basis of the percent of correct answers. More recently Bucking-

[6]Judd, Charles H., Measuring the Work of the Public Schools. *Report, Survey Committee on the Cleveland Foundation,* 1916.

[7]Ashbaugh, E. J., Handwriting of Iowa School Children. *University of Iowa, Extension Division, Bulletin No. 15,* March 1916.

[8]Starch, D., The Measurement of Efficiency in Reading, Writing, Spelling, and English. *University of Wisconsin,* 1914.

[9]DeVoss, J C. *Second Annual Report of Bureau of Educational Measurements and Standards.* Kansas State Normal School, Emporia, Kansas.

[10]Freeman, F. N., *Fourteenth Yearbook of this Society, Part I,* 1915. See also the *Sixteenth Yearbook, Part I,* 1917, Ch. IV.

[11]*Report of a Survey of the Schools of Salt Lake City, Utah,* (1915).

[12]*Report of a Survey of the Schools of Butte, Montana,* Ch. IV (1914).

[13]Witham, E. C., All the Elements of Handwriting Measured. *Educational Administration and Supervision,* 1; 1915, pp. 313-24.

[14]Wilson, G. M., The Handwriting of School Children. *Elementary School Teacher,* 6; 1911, pp. 450-53.

ham has studied the relation between the ability to remember historical facts and the ability to use them. In this study specially devised tests were used. (Ref. 466.)

2. *The Bell and McCollum Test.* This test consists of a series of questions which have been very carefully selected because of their importance. The topics included are: (1) dates-events, (2) men-events, (3) events-men, (4) historic terms, (5) political parties, (6) divisions of history and (7) map-study. The test can be administered in a forty-minute period. (Ref. 270.)

3. *Harlan's Test of Information in American History.* This is a test of historical information based upon the study of Bagley and Rugg, "The Content of American History Texts." Address Chas. L. Harlan, College of Education, University of Minnesota, Minneapolis, Minn.

4. *Starch's American History Tests, Series A.* This test is based upon the facts and principles common to five modern texts. The exercises are in the form of mutilated sentences. Four duplicate forms are available. Address Daniel Starch, University of Wisconsin, Madison, Wis.

VII. Language

1. *Breed and Frostic Scale.* The compositions used by Breed and Frostic in deriving their scale were written by sixth-grade pupils under uniform conditions. A part of a story called *The Picnic* was read to the class, and they were given 20 minutes to complete it. The method of selecting compositions for the scale and determining scale values was similar to that employed by Hillegas. (Ref. 165.)

2. *Courtis Standard Tests in English.* See Reading, below.

3. *Harvard-Newton Composition Scale.* The Harvard-Newton Composition Scale consists of four separate scales, one for each form of discourse; argumentation, description, exposition, and narration. Each of the scales consists of six compositions written by eighth-grade pupils and arranged in order of merit as determined by the marks assigned by teachers, rating them as eighth-grade compositions. For each composition there is given a statement of the most significant merits and defects. Address Harvard University Press, Cambridge, Mass. (Ref. 161.)

STANDARDS. MEDIAN SCORES FOR HARVARD-NEWTON SCALE

Grade	Number of Composition	Median Scores
VIIb	67	60
VIIa	72	64
VIIIb	68	66
VIIIa	61	69
IXb	57	68
IX	61	68

4. Hillegas Scale for the Measurement of the Quality in English Composition for Young People. This consists of ten compositions ranging from an artificial production, whose scale value is zero, to the tenth composition, whose scale value is 9.3. Three of the ten compositions are artificial productions, five were written by high-school pupils, and the remaining two by college freshmen. No two were written on the same topic and they vary greatly in length and type. Each degree of merit is represented by only one composition. (Ref. 172.)

STANDARDS FOR THE HILLEGAS SCALE

Grade	Salt Lake City	Butte	Trabue: Median Score	Trabue: Score above which three-fourths of pupils should rank
IV	2.9	2 34	3.5	3.0
V	3.1	2.87	4.0	3.5
VI	3.8	3.40	5.0	4.0
VII	4.4	3.75	5.0	4.5
VIII	5.4	4.11	5.5	5.0
IX	...	....	6 0	5.5
X	...	....	6.5	6.0
XI	...	....	6.9	6.4
XII	....	...	7.2	6.7

5. Nassau County Supplement. The Nassau County Supplement to the Hillegas Scale consists of nine compositions, seven of which were written by elementary-school pupils on the topic "What I should like to do next Saturday." The compositions of the scale were carefully selected and evaluated by an elaborate method which cannot be even sketched here. Copies may be obtained from the Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 203.)

6. Thorndike's Extension of the Hillegas Scale. This extension is similar to the original scale, except that a larger number of compositions have been used, thereby making a more finely divided scale as well as providing several compositions for each degree of merit in the middle of the scale. Address Bureau of Publications, Teachers College, Columbia University, New York City.

7. The Trabue Completion-Test Language Scales. Trabue has devised a series of Completion-Test Language Scales for the general measurement of language ability. Each scale consists of sentences from which one or more words have been omitted. The position of the omitted words is indicated by a blank. The pupil is to write in the missing words. The relative difficulty of the sentences has been carefully determined, and they have been arranged in order of difficulty. It is claimed for these tests that a pupil's "language ability" is very closely related to his score on these scales. Copies may be obtained from the Bureau of Publications, Teachers College, Columbia University, New York City. (Ref. 202.)

STANDARDS: TRABUE COMPLETION-TEST SCALES

Grade	Median	Grade	Median
II	3.0	VIII	13.3
III	6.0	IX	14.2
IV	8.0	X	15.3
V	9.6	XI	15.8
VI	11.0	XII	16.2
VII	12.3		

8. Willing's Scale. Willing used compositions written by pupils in Grades four to eight on the topic "An Exciting Experience." Several particular exciting experiences were suggested, and 20 minutes was allowed for writing. In determining the compositions to be used for the scale, "all errors in spelling, punctuation, capitalization and grammar were counted and corrected." The compositions selected as samples for the scale were those which had the same rank in "story value" and frequency of errors. Address Bureau of Measurements and Standards, Emporia, Kansas. (Refs. 504, 513.)

For the Denver Survey the following median scores were obtained:

Grade	4th A	5th A	6th A	7th A	8th A
Median	31.5	43.4	50.9	60.2	63.4

9. Buckingham's Grammar Test. In making the survey of the Gary and the Prevocational Schools of New York City, Buckingham used a series of questions upon English grammar. These questions were carefully evaluated upon the basis of difficulty. They have been re-arranged and published by Haggerty. (See below.) (Ref. 466.)

10. Charters' Grammar Test. This test consists of sentences containing an incorrect form. The pupil is to write the sentence in the correct form and give the grammatical reason for doing so. The incorrect grammatical forms were selected from the errors occurring in the oral and written speech of pupils. Address W. W. Charters, University of Illinois, Urbana, Illinois.

11. Haggerty's Grammar Test. Same as Buckingham's, with the questions re-arranged and printed in convenient form. Address Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minnesota.

12. National Business Ability Tests. These include two tests on grammatical form and one on punctuation. In the grammar tests the pupil is to choose between two forms which are given. Address Sherwin Cody, Managing Director, 189 West Madison St., Chicago, Illinois.

13. Starch's Grammatical Scales. Starch has devised three scales (A, B, and C) to measure a pupil's ability to use correctly certain language forms. His Grammatical Scale A consists of a series of exercises arranged in order of increasing difficulty. As tentative standards of attainment Starch gives the following scores for the use of these scales (Ref. 214):

Grade	VII	VIII	IX	X	XI	XII	Freshmen
Score	8.0	8.3	8.6	8.9	9.2	9.5	10.3

14. Starch's Punctuation Scale. Starch has also devised a Punctuation Scale which is similar in form to the Grammatical Scales. The exercises consist of sentences to be punctuated. The following are tentative standard scores of attainment for the ends of the respective school years:

	Grades		High School				University
Year	7	8	1	2	3	4	8
Score	8.0	8 3	8.6	8.9	9.2	9.5	10.3

15. Starch's Grammatical Tests. Starch has also devised three tests for measuring directly a pupil's ability to recognize certain language forms. In Test 1 the pupil is asked to mark the part of speech of each word in a certain printed text. His score is the number he designates correctly in three minutes. Test 2 calls for the

designation of the case of the nouns in another printed test. Test 3 has to do with the tense and mode of verbs.

STANDARDS: STARCH'S GRAMMATICAL TESTS

	Grades		High School				University
Year	7	8	1	2	3	4	8
Score, Test 1	30	33	36	40	43	46	60
Score, Test 2	13	16	20	23	26	30	45
Score, Test 3	13	16	20	23	26	30	45

Address Daniel Starch, University of Wisconsin, Madison, Wisconsin.

16. Thompson's Research Test in Grammar. This is a test of the pupil's ability to indicate the part of speech in a list of words. The feature of the test is a mechanical device for scoring the papers. Address T. E. Thompson, Monrovia, California.

17. Boston Copying Test. This test was devised to measure the ability of pupils to copy printed matter. In giving the test, each pupil was provided with a printed selection which he was asked to copy with pen and ink. In marking the papers the following errors were noted: in spelling, capitalization, punctuation, undotted i's, uncrossed t's; in omitting words, in adding words, in wrong words used, and in misplaced words. (Ref. 164.)

The errors noted consisted of nine different kinds, and the number of each kind made in this test by 4494 pupils is shown by the following tabulation:—

Spelling	5,829
Capitalization	644
Omitted words	4,077
Added words	606
Wrong words used	840
Misplaced words	105
Punctuation	5,876
Undotted i's	8,794
Uncrossed t's	606
Total	27,377
Average errors per pupil	5.54

VIII. Music

1. Seashore's Musical Talent Chart. This chart is based upon the analysis of musical ability and offers a graphic means of representing the pupil's musical ability. Address Carl E. Seashore, University of Iowa, Iowa City, Iowa.

IX. Silent Reading

1. Brown's Silent Reading Test. This test consists of a very interesting reading selection, which is used in Grades III to VIII. Duplicate selections of equivalent difficulty are obtainable. The directions require that the children read the selection silently for exactly one minute, then draw a line around the word which they have reached when the examiner calls "Stop." The number of words read makes the score in speed.

The children are then asked to write as much as they can remember of what they have read. A key is provided for the examiner to use in scoring the papers. On it are listed all the separate ideas contained in the selection. By comparing the child's papers with the key, the examiner determines how many different points there are in what the child read. Then his reproduction is examined carefully to determine (1) quantity and (2) quality of comprehension. Address, Bureau of Research, 25 Capitol St., Concord, New Hampshire. (Refs. 286, 287, 288.)

STANDARDS: TENTATIVE SCORES WITH THE BROWN SILENT READING TEST

	Words per Second	Comprehension	Reading Efficiency
Grade III	3 32	46	127.8
Grade IV	3 55	65	217.1
Grade V	4 40	61	291.0
Grade VI	4 54	68	295 0
Grade VII	4.65	78	322 3
Grade VIII	4.84	79	323.6

2. Courtis Standard Research Tests in English. These are a series of tests devised by Courtis to measure speed and comprehension of silent reading. The series of tests was so complex that the marking of the test papers was a laborious task. For this reason the publication of the tests has been discontinued. (Ref. 291.)

3. Courtis Research Tests Silent Reading (Series R, Test 2). This test is suitable for Grades I to VI. It measures a phase of read-

ing ability acquired in Grades II, III, and IV. In the first part of the test the children read a simple child's story under normal conditions. From this work a measure of the rate of reading is derived. In the second part of the test the paragraphs of the story are reprinted and under each are given five simple questions about the paragraphs. The questions may be answered by "yes" or "no" and they are designed to measure a child's comprehension of the relation existing between the essential elements of the story. The tests are available in two forms of nearly equal difficulty, so that measurements may be made at the beginning and end of the year. Address S. A. Courtis, 82 Eliot Street, Detroit, Michigan. (Ref. 293.)

COURTIS SILENT READING: SERIES R, TEST 2
(Median Grade Scores at the End of the Year)

Grade	II	III	IV	V	VI
Words read per minute	84	113	145	168	191
Questions answered in 5 minutes	16	24	30	37	40
Index of comprehension	59	78	89	93	95

4. Fordyce's Scale for Measuring the Achievements in Reading. This scale consists of a selection to be read, upon which the pupils are required to answer certain questions that have been weighted for determining the comprehension score. The speed of reading is found by having the pupils mark the word they have reached at the end of the stated interval. In order that all may have the information necessary for answering the questions, the pupils are then directed to finish the story. Address the University Publishing Co., Lincoln, Nebraska.

STANDARDS IN PERCENTS: FORDYCE SILENT READING TEST

Test No. 1, designed for Grades III, IV and V.			
Grade	III	IV	V
Speed	90	95	100
Quality	57	71	74
Test No. 2, designed for Grades VI, VII and VIII.			
Grade	VI	VII	VIII
Speed	90	100	100
Quality	41	45	50

5. Gray's Silent Reading Tests. These tests consist of three selections, one for Grades II and III, one for Grades IV, V, and VI, and another for Grades VII and VIII. The selections are so arranged on the pages that the time required to read one hundred words can be readily ascertained. Only one child is tested at a time.

After completing the reading, the child, if in the second or third grade, tells the story to the examiner, who writes it down. In the grades above the third, the child writes all he can remember of the story, and then writes answers to a set of questions which is furnished him by the examiner. The child's score for quality of reading is assigned on the basis of two factors, reproduction and accuracy. Reproduction is determined by the number of words which remain in the child's composition, after all wrong or irrelevant statements and repetitions are stricken out. Accuracy is determined on the basis of ten points for each correct answer. The quality mark is the average of these two. Address William S. Gray, School of Education, University of Chicago, Chicago, Illinois. (Refs. 297, 300a.)

STANDARD SCORES FOR GRAY'S SILENT READING TESTS

Grade	II	III	IV	V	VI	VII	VIII
Rate (words per second)	1.50	2.30	2.20	2.57	2.79	2.69	2.87
Quality	32	37	29	32	39	22	27

6. Haggerty's Visual Vocabulary Tests. The tests prepared by Haggerty, of the Bureau of Cooperative Research, School of Education, University of Minnesota, are but a slight modification of the Thorndike Visual Vocabulary Scales, with the addition of an oral test for children of Grades I and II. This test will be described under the heading of "Oral Reading." Scale R2, of which there is one sheet for children of Grades III and IV and another sheet containing part of the same words and additional more difficult words for Grades V, VI, VII and VIII, is devised in exactly the same way as the Thorndike scales. Methods of scoring are somewhat more simple, and the lists are briefer than those used by Thorndike. Forms of equivalent difficulty are obtainable. Address Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minnesota. (Ref. 301.)

7. Kansas Silent Reading Tests. These tests were devised by F. J. Kelly. Both speed and comprehension of reading are combined in a single mark. These tests consist of graded lists of exercises which have been carefully evaluated. Each exercise consists of the directions for doing something, which is very simple after the pupil has fully understood the directions. His comprehension

of the exercise is measured by what he does. Test I is for Grades III, IV, and V; Test II for Grades VI, VII and VIII; Test III for Grades IX, X, XI, and XII. Address Bureau of Educational Measurements and Standards, Emporia, Kansas. (Refs. 314, 315.)

STANDARDS: MEDIAN SCORES, KANSAS SILENT READING TESTS
(Based upon more than 100,000 Scores)

Grade.....	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Twenty-five percentile . .	2.5	6.1	9.4	9.4	11.8	13.7	16 0	17.9	18.7	22.3
Median Score	5.3	9.5	13.2	13.9	16.2	19.2	22.9	25.6	26.5	29.7
Seventy-five percentile . .	8.2	13.6	17.5	19.8	21.9	26.4	30.4	31.9	33.1	34.1

8. The Minnesota Scale Beta. This is slightly modified form of Thorndike's Scale Alpha (*q. v.*) It is printed in a form which is more convenient for use. This scale is issued in two forms which are approximately equal in value. Address Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minnesota.

9. Monroe's Standardized Tests in Silent Reading. In these tests those features of the Kansas Silent Reading Tests which have proved satisfactory have been incorporated. The exercises have been secured from school readers and other books which children read. Test I is for Grades III, IV and V; Test II for Grades VI, VII, VIII; Test III for Grades IX, X, XI, and XII. These tests are issued in three forms which are equivalent in value. Address Bureau of Educational Measurements and Standards, Emporia, Kansas.

10. Starch's Silent Reading Tests. These tests are similar to Brown's Silent Reading Tests described above, except that different selections are used for the different grades. Address Daniel Starch, University of Wisconsin, Madison, Wisconsin.

STANDARDS: MEDIAN SCORES IN STARCH SILENT READING TESTS
(Attained at the Close of the Respective Years)

Grade or Years.....	I	II	III	IV	V	VI	VII	VIII
Speed (words per second) . . .	1.5	1.8	2.1	2.4	2.8	3.2	3.6	4.0
Comprehension (words written)	15	20	24	28	33	38	45	50

11. Starch's English Vocabulary Tests. These tests are lists of one hundred words, each selected at random from a dictionary.

The child is asked to check the words, the meaning of which he is certain, and to write the meaning after the words of which he is in doubt. The score is the percent of words thus checked or correctly defined. These tests measure the extent of the pupil's vocabulary, regardless of the value of the words. Address Daniel Starch, University of Wisconsin, Madison, Wis.

The following are tentative standard scores for the various years as determined from tests made in four schools (Ref. 333):

	Elementary				High School				University			
Years	4	5	6	7	8	1	2	3	4	1	2	3	4
Scores	30	33	36	39	42	45	47.2	50	53	56	58.5	61	63

12. Thorndike's Visual Vocabulary Scales. Thorndike is the author of three visual vocabulary scales: Scale A, Scale A2 and Scale B. The latter two represent extensions of the former, and were derived by the same method. Scale A2 and Scale B are intended for use alternately or interchangeably, and each is issued in three forms. Each scale consists of graded lists of words, the meaning of which the pupil is to indicate by assigning the words to certain classes. Address Bureau of Publications, Teachers College, Columbia University, New York City.

Standards: No standards have as yet been derived by the use of the Thorndike Scale A2 or Scale B with large numbers of public school children. In the following table the standards of achievement by the use of the Thorndike Scale A (with which the values on Scales A2 and B are supposed to be identical) are given, and serve as tentative standards for purposes of comparison. The score values were obtained by the measurement of the pupils in 18 cities in Indiana. (Refs. 336, 342.)

MEDIAN SCORES IN VISUAL VOCABULARY BY THE THORNDIKE SCALE A

Grades	III	IV	V	VI	VII	VIII
Median Score	4 00	5.26	6 00	6 66	7.29	7.91
Number of Children	1650	2095	2028	1860	1625	1313

13. Thorndike's Scale Alpha and Alpha 2 for Measuring the Understanding of Sentences. Scale Alpha 2 is a slightly more elaborate edition of Alpha. Each scale consists of a carefully graded

series of paragraphs. Each paragraph is followed by several questions which the child is to answer as he reads the paragraph. The pupil's answers to the questions determine the measure of his comprehension. In the *Teachers College Record* which describes the derivation of this scale, there is given a score card for marking the answers to these questions. Address Bureau of Publications, Teachers College, Columbia University, New York City. (Refs. 340, 343.)

The following table gives the median scores for the pupils in 18 cities in Indiana, as reported by Haggerty.

MEDIAN SCORES IN UNDERSTANDING OF SENTENCES BY THE THORNDIKE SCALE ALPHA

Grades	III	IV	V	VI	VII	VIII
Median	5.48	6.56	7.56	8.46	8.72	9.00
Number of pupils	1650	2095	2028	1860	1625	1313

X. Oral Reading

1. Gray's Oral Reading Test. This test consists of twelve paragraphs, arranged in order of increasing difficulty. The relative difficulties have been established experimentally. The child's oral reading of each paragraph is checked for time, and for each of six types of errors: gross errors, minor errors, omissions, substitutions, insertions, and repetitions. Address William S. Gray, School of Education, University of Chicago, Chicago, Illinois. (Ref. 297.)

2. Haggerty's Visual Vocabulary Tests. For Grades I and II these consist of two sheets, one of eight words and the other of phonetic words selected from the Jones test. The words on either sheet are grouped into lists according to difficulty. This difficulty was determined by trial with several hundred primary children. A value is attached to each word according to its ascertained difficulty. The child is asked to pronounce the words aloud; his score is the value attached to the most difficult list of which he can pronounce four out of five words correctly. Two equivalent forms are available. Address Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minnesota. (Ref. 301.)

3. Jones' Visual Vocabulary Tests. Selecting ten of the most widely used primers, Jones found the frequency of occurrence in all the primers of each word occurring in any of them. He used this frequency as a measure of the value of each word. Using the

values thus determined for each word, lists of words were made up as tests. The score is the sum of the values attached to the words which the pupil can pronounce correctly. Address R. G. Jones, Cleveland, Ohio. (Ref. 308.)

4. Price's Oral Reading Tests. These consist of a series of suitable oral reading exercises for Grades II to VIII, inclusive. Two forms are available for each grade. Pupils are scored for words mispronounced, words inserted, words omitted, words transposed and number of words read. Address Supt. E. D. Price, Enid, Oklahoma.

XI. Spelling

1. Ayres' Spelling Scale. This scale consists of a list of the one thousand most frequently used words of the English language. These were determined by means of careful analysis of written material, ranging from friendship letters to some of our best prose. Later the words were classified according to frequency of misspelling for each of the several grades, and the percent of correct spellings for each grade was printed at the head of each list. These standards are for the words when used in dictated lists and without regard to whether the words have been taught in the respective grades or not.

Strictly speaking, Ayres' Spelling Scale is not a scale or test, but a list of words from which tests can be made to measure the ability of pupils to spell the foundation words of the English language. The next four tests were constructed by using words from Ayres' list. Address Russell Sage Foundation, New York City. (Ref. 352.)

2. Courtis Standard Research Tests in Spelling. In these tests words chosen from suitable columns of Ayres' Scale are embedded in sentences, and the sentences are arranged so that they can be dictated at specified rates, which correspond to the rate of writing in the several grades. Each test includes 20 words. The standards set by Courtis are slightly lower than Ayres' Standards for the same words when dictated as isolated words. Address S. A. Courtis, 82 Eliot St., Detroit, Michigan.

3. The Iowa Dictation Exercise and Spelling Tests. These tests, prepared by E. J. Ashbaugh, of the Extension Division, University of Iowa, consist of twenty words embedded in sentences and

an equal number to be dictated separately. Each of the sentences has been constructed so that it is to be written in 30 seconds. Test I is for Grades III and IV, Test II is for Grades V and VI and Test III is for Grades VII and VIII. Address E. J. Ashbaugh, Iowa City, Iowa.

4. The Nebraska Spelling Test. This test, prepared by Dean Chas. Fordyce, of the University of Nebraska, consists of twenty words embedded in sentences which are to be dictated at rates specified for the several grades. The same words are used for all grades. Address Dean Chas. Fordyce, Lincoln, Nebraska.

5. Monroe's Timed Sentence Spelling Tests. These tests differ from the foregoing in that 50 words are used and no test words occur at the end of a sentence. This last feature protects the slow writer. Test I is for Grades III and IV, Test II is for Grades V and VI and Test III is for Grades VII and VIII and for the high school. The normal rate of writing was determined by measuring the speed of 6,000 Kansas school children. Address Bureau of Educational Measurements and Standards, Emporia, Kansas.

6. Boston Minimum Spelling Lists. These consist of a list (for each grade) of "commonly used but often misspelled words." These words have been standardized for the grades in which they are to be taught, and hence constitute lists from which words for testing may be selected. Address Department of Educational Investigation and Measurement, Boston, Mass. (Ref. 355.)

7. Buckingham's Spelling Scale. Starting with a list of about 5,000 words common to at least two out of five spelling books, Buckingham by means of an elaborate statistical procedure, selected two lists of 25 words each. The purpose of the selection was to secure "words which were easy enough in the third grade and hard enough in the eighth grade to afford a test in those and therefore intermediate grades, and which showed regular increases in percent correct from grade to grade." The difficulty of each word was determined in terms of a common unit. Since the difficulty of each word is known, the entire list, or any desired portion of it, may be used as a test. Address Bureau of Publications, Teachers College, Columbia, University, New York City. (Ref. 358.)

8. Jones' Concrete Investigation of the Material of English Spelling. This bulletin presents the results of an investigation to determine "what words, 'grade for grade,' do children use in their own free written speech, and which, therefore, they need to know how to spell." This list has been used as the basis for the construction of tests. One such test has been devised by W. W. Phelan, University of Oklahoma, Norman, Okla., and used extensively in that state. For the *Bulletin,* address University of South Dakota, Aberdeen, South Dakota. (Ref. 370.)

9. National Business Ability Tests. The elementary test consists of 50 words chosen from Ayres' list of 542 obtained from the examination of two thousand letters. The advanced spelling test consists of a list of 50 words which are printed incorrectly. In ten minutes the pupil is to write the words correctly. Address Sherwin Cody, 189 West Madison St., Chicago, Illinois.

10. Rice's Spelling Test. This test has a very great historical importance because it was Rice's report on spelling at the meeting of the Department of Superintendence in 1897 that marks the beginning of the modern movement for scientific measurement in education. (Ref. 378.)

11. Starch's Spelling Scales. These scales have a function which differs from that of Ayres' Scale or a test made from Ayres' list. The latter measures how well pupils can spell the most commonly used words of the English language, while Starch's test measures the size of one's spelling vocabulary. The tests consist of words selected at random from the non-technical words of the English language, with no regard to the frequency with which they are used. Address Daniel Starch, University of Wisconsin, Madison, Wisconsin.

Standards: Starch gives the following standards for his tests based on their use with over 2,500 pupils.

Grade	I	II	III	IV	V	VI	VII	VIII
Percent of words spelled correctly	10	30	40	51	61	71	78	85

These standards are interpreted thus: the average eighth-grade pupil should be able to spell correctly 85 percent of the non-tech-

nical words of the English language, or 85 of the 100 words in any one of Starch's tests. (Ref. 384.)

Tests for Use in the High School

Certain tests of those described in the preceding pages are intended to be used in the high school, as well as in the elementary school. These are Test III of the Kansas Silent Reading Tests, the Thorndike Scale Alpha for Measuring the Understanding of Sentences, Starch's English Vocabulary Test, certain composition scales, the copying test, the Trabue Completion-Test Language Scales, and Starch's Grammatical Tests. For the description of these tests and the use of them, the reader is referred to the preceding pages.

In addition to these tests, many of the others have been applied to high-school pupils. For example, the Courtis Standard Research Tests in Arithmetic, Series B, have frequently been given to high-school pupils, although many of them were not studying arithmetic. However, in applying such tests to high-school pupils it should be remembered that the tests were not designed for that purpose, and it may be expected that they will not be as satisfactory as when used in the way intended.

I. Algebra

1. Coleman's Scale for Testing Ability in Algebra. This test consists of a series of exercises arranged in order of difficulty. Address Supt. W. H. Coleman, Bertrand, Nebraska.

2. Hotze's First-Year Algebra Scales. These scales include tests on the following topics: (1) addition and subtraction, (2) multiplication and division, (3) equations and formulas, (4) graphs, (5) problems. They were restricted to these topics because the author felt "that the main business of the work in first-year algebra was to teach students how to solve typical algebra problems through the use of algebraic symbols." Address Bureau of Publications, Teachers College, Columbia University, New York City.

3. Indiana Algebra Tests. Monroe's Standard Research Tests in Algebra, described below, were incorporated in this series. The other six tests of the series were devised by H. G. Childs, of the University of Indiana. (Ref. 58.)

4. Monroe's Standard Research Tests in Algebra. These consist of a series of six tests. Each of the first five tests is designed to measure the ability to do one of the operations occurring in the solution of simple equations. Address Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas. (Refs. 60, 61.)

STANDARDS: MEDIAN SCORES FOR MONROE'S STANDARD RESEARCH TESTS IN ALGEBRA

Test	I	II	III	IV	V	VI
Number of pupils	2077	1993	2107	2127	2198	1992
Speed, number of examples attempted	14.6	5.4	11.5	10.2	11.2	8.3
Accuracy, per cent of examples correct	96	41	100	94	77	32

5. Stromquist's Preliminary Algebra Tests. This series of tests includes tests upon the following operations: (1) addition, (2) subtraction, (3) multiplication, (4) division and (5) factoring. Address, University of Wyoming, Laramie, Wyoming.

6. Rugg and Clark Standardized Tests in First-Year Algebra. This series includes sixteen tests which are intended to measure all of the types of exercises in the work of the first year. Address H. O. Rugg, University of Chicago, Chicago, Ill. (Ref. 66.)

TENTATIVE STANDARDS FOR RUGG & CLARK FIRST-YEAR ALGEBRA TESTS

(Approximate Median Number of Problems Attempted per Minute for Each Test)

Test	2	5	6	7	8	10	11	12	13	14	15
Most efficient school	4.2	14.5	6.3	13 4	4.6	3.0	4.2	3.5	1.1	7.3	1.4
Average of upper third of 27 schools	3.5	11.6	5.8	12.5	4.1	1.5	3.2	2.9	1.0	5.8	2.5
Ninth school	3.2	10.7	5.5	11.5	3.9	1.2	2.8	2.6	0.8	5.2	3.9
Average of 27 schools	3.0	10.4	4.9	11.2	3.6	1.1	2.7	2.4	0.8	4.9	3.1
Poorest school	2.2	6.5	2.8	7.9	2.4	0.5	1.5	1.4	0.5	2.9	

(Approximate Median Number of Problems Right.)											
Most efficient school	3.2	13.2	5.4	11.8	4.1	1.4	2.8	2.8	0.9	5.3	2.0
Average of upper third of 27 schools	2.7	11.0	4.8	10.7	3.5	0.8	1.6	2.0	0.7	3.9	4.2
Ninth school	3.2	10.0	4.4	9.0	2.9	0.5	1.2	1.7	0.6	3.4	5.9
Average of 27 schools	2.2	9.7	3.8	9.2	2.9	0.5	1.1	1.4	0.6	2.8	7.9
Poorest school	1.2	6.1	2.2	4.3	1.7	0.2	0.1	0.5	0.3	0.1	

Standards for Tests 1, 3, 4, 9 and 16 will be sent to cooperating schools during 1917.

Score for Test 15 is minutes required to solve one problem.

7. Thorndike's Algebra Test. This is a series of eight exercises arranged in order of increasing difficulty as determined by the opinion of competent judges.

II. Drawing

1. Rugg's Scale for Measuring Freehand Lettering for Use in Secondary Schools and Colleges. It consists of a series of 8 samples of freehand lettering, arranged in the order of increasing merit. It may be used in measuring the efficiency of a student's work in freehand lettering. Address H. O. Rugg, School of Education, University of Chicago, Chicago, Ill. (Ref. 156.)

III. Foreign Language

1. Brown's Connected-Latin Test. This test consists of a connected passage of Latin, to be interpreted in terms of its thought content. The pupils are given a specified amount of time in which to interpret and write in English as much of the passage as possible. The translation is scored by means of a key.

2. Brown's Latin-Sentence Test. This consists of a series of Latin sentences ranging from very easy to very difficult. The sentences are graded and evaluated, and each is assigned a scale value.

3. Brown's Formal Latin-Vocabulary Test. A list of fifty isolated words which have been graded and evaluated and a scale value assigned to each. The pupils are scored on their ability to give correct meanings for the words.

4. Brown's Functional Latin-Vocabulary Test. A list of words in the Latin-Sentence Test. The pupils are scored on their ability to re-act to these words correctly in their functional relationships in sentences.

5. Brown's Formal Latin-Grammar Test. This test is made up of twenty constructions in Latin sentences. The constructions are in italics and the pupils are required to name and describe them, but not to translate the sentences.

6. Brown's Functional Latin-Grammar Test. A series of Latin constructions chosen from the Latin-Sentence Test. The pupils are graded on their ability to react correctly to these constructions in their normal settings.

With the foregoing six tests a survey was made of the work in Latin in New Hampshire secondary schools. The high schools of every city and every large town in the state were tested. Most of the private academies and seminaries were also reached. The results are to be set forth in a document entitled *A Study of Ability in Latin.* This will consist of about 300 pages when printed and will be published by the Bureau of Educational Research of the New Hampshire Department of Public Instruction, Concord. For further information address President H. A. Brown, State Normal School, Oshkosh, Wisconsin. (Ref. 277.)

7. Hanus' Latin Tests. These consist of four tests for vocabulary, a translation test, and a grammar test. All of these tests are based on Caesar and Cicero. No words appear in the vocabulary tests "which occur less than one hundred times in Caesar and Cicero." The translation test "contains only constructions which are found at least five hundred times in Caesar and Cicero." The grammar test is based on the sentences to be translated. Address Paul Hanus, Harvard University, Cambridge, Massachusetts. (Ref. 278.)

8. Henmon's Latin Tests. These consist of (1) an easy hundred-word vocabulary test—50 in English and 50 in Latin—containing the words that are common to four widely used first-year books, (2) a standard vocabulary test of 239 words representing all the words common to 13 first-year books and to Caesar, Cicero, and Virgil, (3) a Latin-sentence test consisting of 30 sentences constructed by using none but the 239 words of the Standard Vocabulary Test. Address V. A. C. Hennon, University of Wisconsin, Madison, Wis.

9. Starch's French Vocabulary and Reading Tests. The vocabulary test consists of 100 French words selected at random from a French dictionary. The English equivalents of these words are given on the test sheet, and the pupil is tested by means of the number of English equivalents he can correctly associate with the French words. The reading test consists of simple sentences to be translated.

10. Starch's German Vocabulary and Reading Tests. These are similar to the tests for French. Copies of these tests may be obtained from Daniel Starch, University of Wisconsin, Madison, Wisconsin.

11. Whipple's German Vocabulary Test. This is a vocabulary test to measure the ability of graduate students to read scientific German in the field of educational psychology. A score of 60 has seemed to indicate sufficient ability to read psychology and education in German. Address Guy M. Whipple, Carnegie Institute of Technology, Pittsburgh, Pa.

IV. Geometry

1. Minnick's Geometry Tests. This series of tests is based on the assumption that the demonstration of a geometrical theorem involves the following abilities: (1) the ability to draw the figure, (2) the ability to state the hypothesis and conclusion, (3) the ability to recall facts concerning the figure, (4) the ability to select and organize facts so as to produce the proof. Address J. H. Minnick, University of Pennsylvania, Philadelphia, Pennsylvania.

2. Rogers' Mathematical Tests. These are a series of tests designed to measure several types of reasoning ability in the field of mathematics. The series includes tests on arithmetic and algebra as well as geometry. Address Bureau of Publications, Teachers College, Columbia University, New York City.

3. Stockard and Bell's Geometry Test. This test consists of 70 questions arranged in 20 groups. "These groups involve drawing figures, naming figures, indicating order of development in demonstration, completing statements, stating of converse, definitions, regular polygons, parts of a demonstration, angular relations, area of trapezoid, angles in polygons, angles in circles, congruency of triangles, similarity of triangles, loci, auxiliary lines, simple constructions, ratio and proportion, algebraic expression of geometrical relations, and equivalent construction. The questions are asked in such a way that many pupils are able to complete the list in forty minutes." (Ref. 223.)

V. History

1. Sackett's Scale in Ancient History. This scale consists of a series of eight tests, or exercises, the questions of which call for the essential information in the field of ancient history. The items of this information were determined upon the basis of a careful examination of an American history text and the judgment of

experts in the field. Address L. W. Sackett, University of Texas, Austin, Texas. (Ref. 273.)

VI. Physical Training

1. Rapeer's Scale for Measuring Physical Education. This is a score card for judging five aspects of the results of physical education, *i.e.*, health, physiological efficiency, physical development, physical ability and mental qualities. Address L. W. Rapeer, San Juan, Porto Rico.

VII. Physics

1. Starch's Tests in Physics. The tests consist of 75 sentences from which certain words have been omitted. The sentences and the words to be omitted have been so chosen that a pupil cannot supply the correct words unless he knows certain physical facts or principles. The facts, principles, and laws upon which these sentences are based were determined by examining five widely used textbooks. The 102 facts, principles, or laws which were treated by all five of the books are the ones which the pupils must know to do the tests correctly. Address Daniel Starch, University of Wisconsin, Madison, Wisconsin.

Distributing Centers

Information in regard to particular tests can always be obtained by writing directly to the authors. Very many of the tests and scales, however, can also be secured from various distributing centers, principally University Bureaus of Cooperative Research. For the convenience of those interested, the addresses of the six most important distributing agencies are given below:

1. Bureau of Publications, Teachers College, Columbia University, New York City.
   Tests by Thorndike, Stone, Hillegas, Trabue, Woódy, Bonser and many others.
2. Russell Sage Foundation, Division of Education, New York City.
   Ayres Writing and Spelling Scales.

3. S. A. Courtis, 82 Eliot Street, Detroit, Michigan.
   Research Tests in Arithmetic (Series A and B), Reading (Series R), Writing, (Series W), Spelling, (Series S).
4. School of Education, University of Chicago, Chicago, Illinois.
   Gray's Reading Tests, Cleveland Arithmetic Tests, Rugg's Algebra Tests.
5. University Supply Association, Madison, Wisconsin.
   Tests by Starch and Henmon.
6. Bureau of Educational Tests and Standards, Kansas State Normal School, Emporia, Kansas.
   Tests by Monroe, Kansas Silent Reading Tests.
7. Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minnesota. Tests by Haggerty.

Very often schoolmen who undertake measurement work without previous training meet difficulties in the scoring and tabulating of the results, or in interpreting them, which they are unable to overcome without assistance. Ordinarily, this assistance can be had for the asking from any worker in the field, but particularly from Departments of Education in State Universities. A letter, or better a personal visit, will often not only clear up in a few minutes a misunderstanding that otherwise might have led to failure and discouragement, but also bring into mutually helpful relations the so-called theoretical and practical educational forces of the state. The supreme service of educational testing is that it reveals problems, stimulates attempts at solution, and affords measures of the success of the efforts made, but even the simplest educational problems are so important, so complex, that they demand the united cooperative efforts of educational workers of every type.

CHAPTER VIII

RELATED FORMS OF EDUCATIONAL INVESTIGATION

W. A. AVERILL
State Education Department, Albany, N. Y.

The broadest interpretation of the title of this chapter includes everything in the whole school world which remains over and above the measurements of classroom work discussed in the other chapters of this Yearbook. It would call for the consideration of a formidable array of topics of all sorts—from the dust content of schoolroom air to the nature and results of educational legislation—a range quite beyond the possibilities of this part of the Yearbook. Accordingly, the aim of this chapter will be two-fold; (1) to call the attention of superintendents to the different types of investigations being made, and (2) where investigations are of a conventional type, as cost accounting, study of enrolment, etc., to indicate present tendencies, rather than to give a detailed statement of the entire field.

Content of the Course of Study. Possibly the most significant type of investigation, as well as one immediately related to classroom measurement, is the analysis of the curriculum, both in elementary and high schools, with special reference to the content of the courses of study and to the relative amount of time devoted to the different subjects. Ayres' investigations in spelling have pointed the way to the elimination of waste in accomplishing given results and have shown which words among the thousands taught were proper tasks for the children of the various grades. Similarly, the investigations of Bagley and Rugg in history,[1] Jessup in arithmetic, and others of the same type have served to indicate the methods by which a scientific curriculum may eventually be constructed.

Pupils' Marks and Scholarship Ratings. Another type of investigation most intimately related to classroom measurement is the

[1]For these and other references consult the bibliography, Chapter XIII.

study of the marks given pupils by teachers. Two growing tendencies are worthy of mention; first, superintendents are recording and studying the marks received by a class in order to discover what percentage of the pupils receive superior, satisfactory and poor ratings, and they are rating teachers, and making adjustments of work, on the basis of comparative percentages of failure; second, they are accumulating such data from year to year and subjecting them to careful statistical and graphical analysis. That is, school men are no longer content to hope and believe that their work is going well; they are attempting systematically to secure exact knowledge of the effects of their efforts by a scientific study of all available data.

Rating the Teacher's Efficiency. Immediately related to the measurement of the work of the pupils is the estimate of the teacher's ability. The tendency appears to get away from general comment and to record specific details about the recitation itself. A rating card of the general type contains such questions as:

Does the teacher show skill in habit formation?
Does the teacher show skill in stimulating thought?
Is the subject matter organized?
Are proper habits of action developed both in and out of the classroom?
Is the teacher interested in the life of the community?

A rating card which goes into particulars contains such questions as:

Number of minutes lost in calling the class?
Number of minutes lost in distributing material?
Extent to which the recitation was confined to the text? Related to the pupils' lives and experiences?
Number of questions which caused the pupils to think before answering? Number requiring merely yes-or-no answers?

Another tendency of teacher rating takes the form of self-examination, of which the following is an illustration:

Clearing up pupils' difficulties.

1. What plan or method have I for discovering the difficulties which the class as a whole may have, and specific difficulties which individuals may have? For providing the particular help a pupil needs to clear up his own difficulty?

2. When a large enrolment makes it practically impossible to provide the special help needed by individuals, what is my method of providing it?
3. Do I appreciate that a pupil may lose a half year or a full year as the result of my failure to give him just the help he needs at the time he most needs it?
4. Am I inclined to shift the responsibility for pupils' failure to causes outside myself?
5. What evidence have I that I am successful in clearing up pupils' difficulties?

Still another important type of investigation connected with the rating of teachers, and one which is rapidly growing in favor, is the study of stenographic reports of actual recitations.

Promotions and Non-Promotions. The question as to the number of pupils promoted at the close of the term arises naturally after mention of the teacher's efficiency. In the limited space of this chapter it is enough to call attention to the fact that superintendents are finding it desirable to record separately the following types of promotion:

1. The mid-term promotion, where the pupil has spent about half the regular time in each of two grades.
2. The double promotion, in which the pupil is made to skip an entire grade.
3. The straight, earned promotion to the next higher grade.
4. The un-earned promotion, meaning that the pupil has not "passed," but has been in one grade so many terms that there is no longer any benefit in keeping him there.
5. The conditional or trial promotion.
6. The failure, or non-promotion.
7. The demotion to a lower grade.

In comparisons of different cities, investigators should have these distinctions in mind, otherwise they may not be comparing the same things. In other words, the tendency in this field, also, is to define more sharply the analysis that is made.

Teachers' Reasons for Pupils' Failures. The causes of non-promotion are a well-known subject of investigation and report, but often the investigation is limited to asking the teachers to write opposite the name of each non-promoted pupil a reason for the failure. Under these conditions teachers will usually ascribe about 40 percent of the failures to the pupils' mentality and 20 percent to

the pupils' lack of application. Here, too, the tendency is to push the analysis further and demand reasons based upon objective evidence.

Elimination. Closely associated with the examination of figures relating to promotions, is the consideration of the extent to which pupils have dropped out of school before promotion time. In the past it has been customary not to count as non-promoted any pupils who have dropped out before examination. This improves the showing, but in line with modern tendencies it is proving more valuable in getting at actual school facts to keep a record of the following types of withdrawal and to give them their full weight in any use which is made of promotion figures.

1. Transfer to another part of the same public-school system.
2. Transfer to schools without the local system.
3. Removal from city with an implied continuance of schooling.
4. Schooling stopped for the current term:
   a. Poor health.
   b. Bona fide cases of "needed at home"
   c. Actual poverty.
5. School given up as a bad job:
   d. Incapacity
   e. Indifference
   f. Disciplinary.

Vocational Education. Early investigations of the causes of withdrawal from school, particularly those which recorded the statements of pupils and parents, as in the Minneapolis survey, brought out the fact that neither pupils nor parents regarded schooling beyond the grades as of very material aid in gaining a livelihood. A few studies were made, correlating wages and salaries with time spent in school. Then the rapid growth of manual training, shop courses, and definite training for the industries, supplied material for a whole literature of vocational investigation and survey. This soon passed beyond the content of the curriculum to the industries themselves, culminating in the thoroughgoing survey of the Cleveland type, in which the whole city is analyzed vocationally and the data, quantitative as well as qualitative, are practically and constructively related to the schools and their program. It is interesting to note that in this connection, the first chapter of the Bloom-

ington, Indiana, survey report treats of the occupations of the women and the men employed in the city.

Ages of Pupils and Progress Through School. It would be presumptious to re-state the principles of age and progress research so clearly outlined by Dr. Leonard P. Ayres in his *Laggards in our Schools* and *The Identification of the Misfit Child.* The nine categories of pupils that result from the possible combinations of three age and three progress factors are appearing year by year in a larger number of superintendents' reports. The tendency to get away from a liberal margin of "years of normal age" which would enable any city to have a comfortable preponderance of "normal" pupils is in keeping with the demands of modern supervision and research. "To obtain for one city a record which will exceed the record of other cities" may serve the purpose of self-congratulation, but the more nearly correct picture of the situation which is obtained by using a single year age-limit for each half-grade (as is now done in New York State) is far more valuable, if less flattering, to the local superintendent. Present practice figures age at the time of *beginning* a given grade or completing it, rather than "being in a grade" some time during the school year, and the preference is for the time of beginning. The actual statistics may be gathered any time during the early autumn. Bachman advances good reasons for figuring ages on August 31, the beginning of the official school year; the superintendents of New York State have chosen September 15 for the age date. The investigations of age, progress, elimination and retardation have become far too numerous to list separately in a chapter of this length. A new type of investigation is, however, to be seen in the occasional attention given in these studies to the exceptionally capable pupils who have been almost forgotten in our efforts in behalf of their more unfortunate classmates.

Cities and Villages Studied in Groups. For the investigation of enrolment by grade, elimination, age and progress through school, the New York State Education Department has divided the cities, villages and union high schools with elementary departments into groups based on the elementary enrolment. By means of mechanical tabulation, the data supplied by several hundred communities

throughout the state are quickly tabulated, analyzed, and returned to the systems which contributed the information. Returns will be made, it is believed, in time to be of use to the superintendents early in the next school term. Beginnings in this form of 'educational accounting' have been made in the principal cities of the state and in 12 cities with an elementary enrolment 3,000 to 5,000; 26 with 1,000 to 3,000; 44 villages with 500 to 1,000; 58 with 300 to 500; 227 enrolling 100 to 300 elementary pupils and about 100 union schools with fewer than 100 elementary pupils. The superintendents of these school systems will be glad to exchange data with *systems of like size* throughout the country.[2]

Superintendents willingly receive comparisons of their systems with others of like size similarly situated. The first reaction toward this matter is an attempt to *explain away* the retardation and other defects; it is only afterwards that the problem of *doing away* with it is attacked. Reducing the figures to common denominators of size, wealth, foreign elements and shifting population wins the confidence and cooperation of superintendents and principals and paves the way for constructive work. The mechanical tabulation of the returns also makes possible the correlation of progress-through-school with such factors as teachers' salaries, principals' salaries, number of pupils per teacher, per capita assessed valuation of the school district, number of hours devoted daily to supervision by the principal, and in small schools, the number of grades taught by one teacher.

Investigations Giving Rise to Permanent Records. It is of interest to follow the usual course of many investigations of the type mentioned thus far in this chapter. Often the first stimulus to action is a questionnaire received by the superintendent; next a survey of classroom work is made with standard tests; then an age-progress survey follows, and the superintendent begins to incorporate into his report some of the outcomes. Soon the giving of standard tests and the statistical study of school records become a part of his regular work. Finally the items of information necessary for these investigations become the subjects of permanent record in the superintendent's office and appear regularly in his report.

[2]Address W. A. Averill, State Education Bldg., Albany, N. Y.

Groups of Related Investigations. There should be mentioned here quite an array of investigations, largely correlated with retardation, progress and the achievements of pupils as measured by standard classroom tests, which for want of space can only be listed.

Studies having to do with

1. Physical conditions of the children in school (hearing, eyesight, development, etc., medical inspection, school lunches, heating, ventilation, school furniture, etc.).
2. Home conditions (sleep, breakfast, play, home study, outside work, etc.).
3. Special classes (blind, deaf, mentally defective, tubercular, super-normal, special promotion plans, coaching rooms, supervised study, etc.).
4. School organization (school programs, recesses, fatigue, vacations, division of school year, kindergarten, junior high schools).
5. Compulsory attendance (delinquency, discipline, truancy).
6. Social and moral welfare.

Organization of City School Systems. Different in type though related are investigations dealing with the organization and administration of city school systems, as represented by the work of Cubberley, Strayer, Thorndike and others. Attention is centered on the functions of the different officers of the system, and the chief purpose is perhaps to determine the proper alignment of the functions of the school board and the superintendent. The conception of a school organization in which the superintendent supervises the educational work and a coordinate business officer manages the finances and school plant, is being abandoned in favor of an organization in which the superintendent, as the chief executive officer of the board, is, next to the board itself, the head of the entire system, and the one to whom all other officers, both business and educational, are subordinate. Investigations in educational legislation show this same trend. Legislation recently passed in New York State gives every city a school organization on the plan just mentioned.

School Buildings. An important series of investigations is that dealing with the subject of school buildings. The surveys made in Cleveland, Oakland and Milwaukee are typical. The grouping of types of school architecture by decades, as in the

Cleveland survey, is a significant modern tendency. Sanitation and hygiene naturally play a large part in these studies, as do ventilation, illumination and safety from fire. Related to this group are occasional investigations of janitor service.

Extension and Social Service. Much has also been said and written about the wider use of school buildings—their use at night and in vacations, school-board policies with reference to permitting various organizations to use school buildings and charging for their use. These surveys, however, lead into the field of community welfare and belong rather to the adult world.

The Cost of Educational Work. Finally, no investigation of educational conditions or work can be considered complete which does not show the cost of obtaining the results achieved. The earlier investigations of cost were largely devoted to teachers' salaries; indeed, for many years salaries and "other expenses" were about all that could be culled from school accounts. Now, investigations of school cost fall into three categories: (1) those dealing with the classification of payments for school purposes, (2) those treating of the accounting procedure which gives the desired classification and (3) studies of comparative cost by items, schools and cities.

The present tendencies may be stated as follows:

The school budget is designed according to function, or the kind of work done or service rendered. The four main functions are (1) regulative and executive service, (2) property (acquisition, construction, equipment, maintenance, and operation), (3) instructional, or supervision and teaching, and (4) extension and social service.

School moneys are appropriated by main functions only, all details of expenditure are left to the board of education.

The classification of expenditure involves four things about every payment made, namely, (1) the function subserved, (2) the character of the payment as a fiscal transaction, (3) the detailed object of the expenditure and (4) the location in the system to which it is chargeable.

The accounting procedure itself has been simplified by abandoning large unwieldy forms in favor of smaller loose-leaf sheets; by using flexible code symbols at the top of columns; by using voucher-checks and warrants in place of the old forms of voucher

jacket; and by the standardization of the size of all blanks and cards to the new dimensions used commercially.

Accounts have been rendered more valuable by including a register of orders, a register of accounts payable and an appropriation ledger showing at any time the unencumbered balance as well as the unexpended balance.

Finally the information obtained is reduced to proper units for comparative analysis—per capita units, for instructional functions and square and cubic foot units for building, maintenance and operation functions. Graphic presentation is employed where ever possible.

Modern Office Methods and Equipment. As a subject for present and future investigation, mention should be made of the equipment of superintendents' and school board offices and "modern business methods" of handling clerical work and routine. The adding machine, the calculating machine and the slide-rule will enable one or two clerks to perform a surprising amount of statistical work. The T-square and ruling pen are easily mastered tools for graphic presentation. Time for research is often to be gained by a reorganization of office routine and the elimination of unnecessary records and procedures. Finally, larger systems can solve many of these problems by resorting to the mechanical tabulation of statistics, both educational and financial.

Conclusion. The breadth of the field presented by educational investigations as a whole has made their mention in this chapter disconnected and their discussion, when possible at all, so brief as to be perhaps inadequate. After all, we must revert to the subject of the other chapters of this Yearbook—the measurement of the educational work itself—as that which is most important, because it is most closely related to the purpose of the schools, the education of the children who attend them. Other types of investigation should serve to throw all possible light on the local system in which given educational results appear. They should be of aid in effecting that interpretation of classroom work which will make for fair criticism and constructive suggestion. In fine, these related forms of educational investigation stand as an interpretative background to the more special measurements that mark the achievements of pupils in the schools.

CHAPTER IX

STATISTICAL TERMS AND METHODS

B. R. BUCKINGHAM
Educational Statistician, State Board of Education, Madison, Wisconsin.

As educational reports multiply, it becomes increasingly evident that there is undesirable variation in both their method and language. A standard procedure and terminology will be helpful, both to makers and to readers of reports. Moreover, the application to education of statistical methods developed in other fields needs to be made clear. It is the purpose of this paper to state what appears to be the best practice and to note the assumptions on which the practice is based.

In certain sciences a higher degree of accuracy is required than is possible with a single measurement. Accordingly, many measurements are made of the same magnitude, and their average taken as the true amount. It was early ascertained that these measures showed a certain definite arrangement, in accordance with which values at or near the average of all the measures were of greatest frequence, and measures which differed from the average were less frequent the greater their difference from it. Further, it was found that the number of measures greater than the average was likely to be equal to the number less than the average by the same amount. The graphical representation of this kind of a series took the form which we have lately become familiar with as the 'normal' or 'probability' surface (see Figure 2). Measures which differed from the average were thought of as being in error, and this lead to the development of what has been called the "theory of error"—a theory involving the symmetrical arrangement of measures of the same thing about the true measure as represented by their average.

The assumption was next made that there was an analogy between a series of measurements of the same thing, and a series

of single measurements of each of a number of individuals alike in some important characteristics. This analogy was found to work out very well for biological data. Economical data, on the other hand, do not appear to afford the same analogy, and since, prior to the recent application of statistics to education these methods have been used mainly by biologists and economists, statisticians were divided into two schools—the one adhering to the doctrine of the theory of error, the other rejecting it. The biologists, broadly speaking, belong to the former and the economists to the latter school.

The evidence is strong that educational measurements tend to resemble those of biology in their structure, *i. e.*, that the theory of error applies. Published results often fail to reveal this, but when such is the case, it is almost always due to scanty data, or to the selection of the children, or to the measuring instrument itself.

It is to be understood that the validity of the statements in the succeeding pages of this chapter rests on the assumption that educational measurements substantially conform to the arrangement which exists among many measurements of the same thing, *i. e.*, that they conform somewhat closely to the theory of error, and exhibit approximately 'normal distribution.'

A *variable* is a quantity which under the conditions imposed may assume different values throughout the discussion. In repeated measurements of the same thing, the obtained measures differ through 'error.' In education, where we make numerous single measurements of different pupils grouped together on the basis of some common characteristic, each different value constitutes a value of the variable. The number of problems solved correctly by seventeen girls in an upper eighth-grade class was as follows: 9, 3, 7, 8, 5, 6, 7, 6, 5, 7, 5, 6, 4, 8, 7, 4, 5. Each different number in this series is a value of the variable "performance of upper eighth-grade girls in solving specified problems in arithmetic."

Attention must be given to the nature of the units in which the variable is expressed. They may indicate *mid-points* or *lower limits*, and they are influenced by the accuracy of the measurements

upon which they are based. In the Nassau County Supplement to the Hillegas scale, the samples have the following values: 0, 1.1, 1.9, 2.8, 3.8, etc. If a person reports *to* the nearest scale value, he is using each value as a mid-point, and his rating of a composition as 1.9 will mean that it is nearer to 1.9 than it is to 1.1 or to 2.8; in other words, his 1.9 will mean from 1.5 to 2.35. If, however, he rates a composition as 1.9 which has at least the merit of the scale sample at 1.9, but not as much merit as the sample at 2.8, he is using the scale-value as a lower limit, and his 1.9 will mean from 1.9 to 2.8. If he reports values between those given on the scale such that a composition slightly less meritorious than the one at 1.9 would be given a value of 1.8 or 1.7, and one slightly more meritorious than the one at 1.9 would be given a value of 2.0 or 2.1, then he is attempting to secure a greater degree of accuracy, and his 1.9 would mean 1.85 to 1.95. Thus, the same rating may have materially different meanings. It is necessary, therefore, that the units be accurately defined and that they be given the same meaning throughout a given discussion.

A *series* in statistics is a grouping of the obtained measures of the variable by steps or classes. According to the nature of the units, series may be either discrete or continuous. A *discrete* series is composed of separate integers. If we record the number of children in each class in a school system, all of our measures will be of this nature. For example, between classes of 20 children and classes of 21 children there can be no measures. A *continuous* series, on the other hand, is one which is capable of an indefinite degree of sub-division. The height of children, their weight, the time they take to do a given task, their general intelligence, their ability in specific ways, all of these when measured yield continuous series. We may represent the height of a child as 60 inches, or his mental age as 10 years. In these cases, however, our measures are only short-hand expressions for ranges whose mid-point is the measure as reported. Persons reported as 60 inches in height may be anywhere between 59.5 and 60.5 inches. Most measurements in education yield continuous series. Even when series, taken at their face value, are discrete, such as the numbers of problems solved, they should be treated as continuous series, because

in the abilities which they are supposed to indicate, no gaps occur. To say that in a certain class six pupils got 10 examples right, only indicates that they had an ability which permitted them to work 10 examples correctly, but not 11. Their abilities might be more accurately represented, if we had material with which to do it, by 10, 10.27, 10.5, 10.84, etc.

Measures when first received are unorganized, and no definite impression may be obtained unless their number is very small. The first procedure is to divide them into classes. For example, 700 measures were obtained from eighth-grade boys who wrote Test No. 15 of a series of "Progressive Spelling Tests." The test consisted of 50 words. Clearly 700 measures in a hap-hazard arrangement convey almost no information as to how well these boys performed. We may classify the measures, as in Table I, by indicating the number of boys who spelled no words correctly, one word correctly, two words correctly, etc. This table consists of 51 classes.

TABLE I

DISTRIBUTION OF 8TH-GRADE BOYS ACCORDING TO THE NUMBER OF WORDS THEY SPELLED CORRECTLY "PROGRESSIVE SPELLING TESTS," NO. 15

No of Words Correct	No. of Pupils	No of Words Correct	No. of Pupils	No. of Words Correct	No. of Pupils
0	7	17	15	34	22
1	14	18	20	35	15
2	9	19	19	36	16
3	6	20	18	37	12
4	11	21	25	38	20
5	13	22	20	39	13
6	18	23	20	40	11
7	11	24	14	41	14
8	11	25	15	42	7
9	13	26	20	43	9
10	9	27	23	44	5
11	13	28	22	45	2
12	8	29	22	46	3
13	17	30	25	47	4
14	12	31	19	48	2
15	19	32	24	49	1
16	14	33	18	50	0
				Total	700

Although Table I brings the measures into an orderly arrangement, it does not show the structure of the series as clearly as does Table II which is derived from it, and which exhibits but 10 classes.

TABLE II
DERIVED FROM TABLE I

No. of Words Correct	No. of Pupils
0– 5	60
6–10	62
11–15	69
16–20	86
21–25	94
26–30	112
31–35	98
36–40	72
41–45	87
46–50	10
Total	700

In this table each group of measures, *e. g.*, 0 to 5, 6 to 10, etc., is called a class.[1] For the Class 6 to 10, 6 is called the *lower limit*, and 10 the *upper limit*. The *class interval* is the difference between the upper and lower limits. In this instance the class interval is 5 words. The number of measures within each class interval is called the *frequency* of the class, and its conventional symbol is "f." The entire tabular arrangement, consisting of classes and their frequencies, is called a *frequency table or frequency distribution.*

On the one hand, it is desirable in the interest of accuracy that the number of classes of a frequency table be large, in order that the intervals may be small. Obviously, it is much more accurate to say that 7 pupils spelled no words correctly, 14 spelled one word correctly, etc. (Table I), than to say that 60 pupils spelled from 0 to 5 words correctly (Table II). On the other hand, it is desirable that the number of classes be small enough to afford a generalized or typical result. A good rule to follow is: arrange the data in as many classes as will secure the greatest regularity of frequencies, *i. e.*, in accordance with which the frequencies will as nearly as possible increase to a maximum near the middle of the range, and then decrease to a corresponding minimum. While the number of classes in Table I, according to this criterion, is too large, it might not be too large if we had five or six thousand measures.

The graphic representation of the frequency table is called a *frequency surface* or *frequency polygon.* Figure 1 is a frequency

[1]There is a slight irregularity in the class intervals in that the first class contains six units.

surface corresponding to Table II. Observe that the measures, grouped by classes, are arranged along the base line. The height of each column erected above a given class corresponds to the frequency of the measures for that class. These frequencies may be read from the vertical scale at the left of the figure.

In Figure 1 the columns rise on either side of the tallest column with an approach to regularity. If, instead of 50 words, we had 500, so that we could make 501 classes, and if the number of pupils, (*i. e.*, of measures) were indefinitely increased, the perimeter of Figure 1 would approach the form of a smooth curve having a single peak near the center and sloping toward the base line to the right and left.

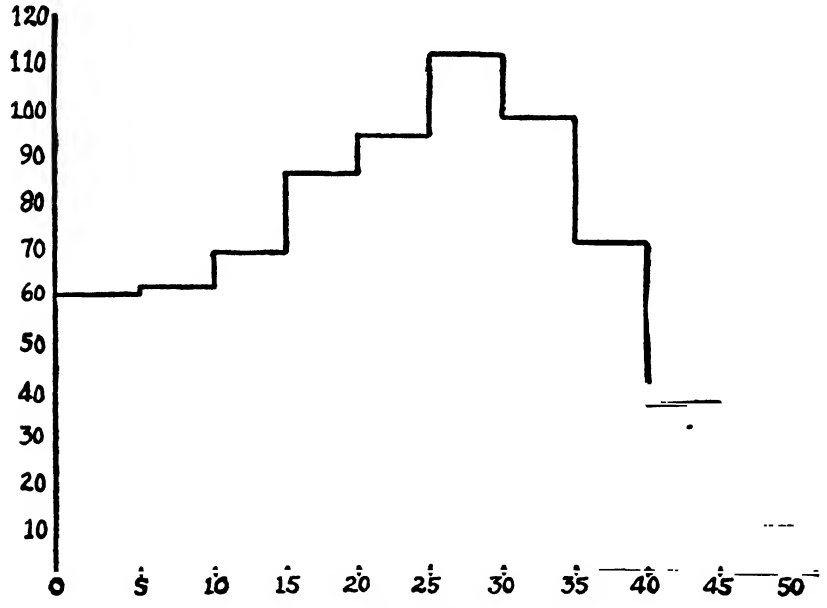

FIGURE 1. Frequency of Correct Spellings

Horizontal distances represent numbers of words correct; vertical distances represent numbers of pupils. Data from Table II.

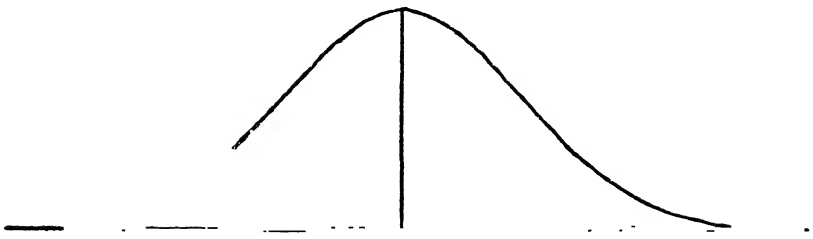

FIGURE 2. Normal Surface of Frequency

The base line of such a curve is called the *abscissa.* Along it are represented the values of the variable. The perpendiculars from the curve of the *abscissa* are called *ordinates.* They represent the frequencies of the values of the variable. The longest perpendicular is called the *maximum ordinate,* and the measure represented on the abscissa at the foot of it is the most frequent one. If a curve is of the symmetrical form shown in Figure 2, the maximum ordinate divides the area between it and the abscissa into two equal parts. A curve of this form is variously called the Gaussian curve, the curve of error, the probability integral, or the normal curve of frequency.

If a smooth curve were drawn to fit as closely as possible the frequency surface of Figure 1, this type form (Fig. 2) would not be accurately reproduced. Spelling ability, however, among eighth-grade boys may, nevertheless, be distributed 'normally.' The present state of our knowledge does not permit us to be sure that the words selected would register all abilities among eighth-grade boys. Again, details, not here reproduced, indicate that the number of children was too small to bring out the real distribution. One small school furnished more than half the children who spelled from 0 to 5 words correctly, and more than one-third of those who spelled 0 to 15 words correctly. This unusually low performance causes the first three columns of Figure 1 to be unduly high. The remedy, of course, is to obtain returns from a great many more children.

Although the arrangement of many measures in a frequency table and their graphic representation by a frequency surface permit the mind to grasp the significance of the data far more readily

than is possible without such devices, nevertheless further simplification is generally desirable. The more numerous the measures and classes of measures, the greater is the need for concentrated information—that is, for a single simple expression which shall contain in itself a summary of the whole series. This is the purpose of averages.

The term *average* has both a general and a specific meaning. In its general sense it means any expression intended to give by a single figure, the general weight or typical measure of a series.[2] In its restricted sense, it refers to the arithmetical mean, *i. e.*, to the sum of the measures divided by the number of them. This is what we generally mean when we speak of *the* average. In educational statistics, the term used to express the general concept of average is *central tendency* (C. T.). As has been stated above, series of educational measurements, when properly derived, tend to exemplify the theory of error, *i. e.* (to state the case very crudely), there is a part of the series where the measures are most numerous, while the other measures become fewer, the more they differ from the most frequent measure. Unless this condition is approximated, it is not appropriate to speak of a 'central tendency.' We may, indeed, make an arithmetical computation. The average of the annual enrolments of an institution which is steadily increasing in size may thus be found and will have its uses, but it will not be typical of any tendency in the measures.

Since we are permitted, however, from the nature of educational data to use the term 'measure of central tendency,' it will be best to use the term 'average' in its commonly accepted sense, as meaning the sum of the measures divided by the number of them. Some writers call this the "arithmetical mean," or simply the "mean."

The three measures of central tendency (C. T.) most frequently used are the average, or arithmetical mean (A), the median (M), and the mode (Z). The median has been defined as the midmost measure of a series whose measures are arranged in order of size, beginning with the smallest or the largest. According to this definition, the median is the $\frac{N+1}{2}$th measure (N being the number

[2]See, for example, Zizek, Franz. *Statistical Averages.*

of measures). There are objections to this conception of the median, and recent workers have defined it as a *point* on the scale on each side of which half the measures lie when they are arranged in order of magnitude. The significance of this definition is that it emphasizes the idea of a *point* on the scale rather than a measure. The measures in Table I are arranged in order of magnitude. Half the number of measures (pupils) is 350. From the beginning through Step 23, there are 342. Eight more measures are needed to complete the 350. These are all contained in Step 24, which may, therefore, be called the "median step." Since the series is to be regarded as continuous, we must seek the M—the "point on the scale on each side of which half the measures lie"—somewhere in Step 24, *i. e.*, between 24 and 25. We now assume that the 14 measures entered at "24" in the table are uniformly distributed between 24 and 25. Since we need 8 of them, M will be 8-14 of the way from 24 towards 25. Hence the M is 24+8-14 or 0.57 *of the step.* The step is 1; therefore the M is 24.57. If we count from the other end of the table, we obtain the same value, thus checking the computation.[3]

If the data were available only as in Table II, the M would be found to be 24.88, instead of the more correct 24.57, as computed from Table I. Thus, the greater regularity of Table II is gained at the usual sacrifice in accuracy.

Series may have but single measures at each value, and there may be more or less wide gaps in the values. If the number of measures is odd, M is the middle one; if even, it is the average of the two middle measures. For example, if the expenditures per pupil for 35 cities are listed in the order of their amounts, M is the expenditure of the 18th city. If there are 36 cities, M is the average expenditures of the 18th and 19th cities.

The median may be used when the items have been merely ranked, instead of measured. Thus a teacher may arrange the compositions of a class in the order of their merit according to her

[3]If there were 701 measures made, let us say, by including one pupil who spelled 37 words correctly, we should find the 350.5th measure. We should then need 8.5 of the 14 measures at Step 24, and M would be 24.61.

judgment.[4] The middle one (or the two middle ones, if there are an even number of them) is the typical performance of the class. Such typical samples may be compared with others similarly obtained at later dates, and thus a showing of the progress of the class may be made. To add to the definiteness of this, the median samples may be measured by a composition scale.

It was pointed out above that the nature of the units in which measures are expressed must be clearly apprehended and consistently adhered to. Failure to do this introduces error in all the results of statistical analysis. An illustration of such an error made in the computation of medians may be found in the report of the Salt Lake City Survey. In the Table (p. 140) showing the distribution of composition scores by grades, failure to take account of the nature of the unit resulted in reporting medians which were too small by an amount greater than the average annual improvement from grade to grade. The teachers rated the compositions according to the original Hillegas scale. The values of the samples on the published scale were 0, 183, 260, 369, 474, 585, 675, 772, 838, and 937. In reporting the ratings of the teachers, however, these values were called respectively 0, 1, 2, 3, 4, etc., and in computing the median these *latter values* were the ones used, and they were called the *mid-points* of their respective class-intervals. Accordingly, when a teacher rated a composition as 3, (really 369) she was regarded as placing it between 2.5 and 3.5. As a matter of fact, she was placing it as nearer to 369 than to 260, or 474, and the class-interval was therefore from 314.5 to 421.5. The median for the 4th grade was found to lie four-tenths of the way into the class called "3" in the table, and, since the lower limit of this class was taken as 2.5 and the interval as 1, the median was reported as 2.9. The lower limit, however, for the class was really 314.5 and the interval 107. This yields a median of 357.3 ($314.5 + 107 \times 0.4$) or, shifting the decimal point for convenience and expressing the result correctly to one decimal place, 3.6, instead of 2.9 as reported.

The mode is the most frequent value in a series. As such it is

[4]This may be done very accurately by making all possible comparisons of two papers and giving the better paper of each pair a preference mark. The paper having the greatest number of preference marks is the best paper; the one having the next greatest is the next best, etc.

the most evident measure of C. T. It is what most people think of when they speak of the average. A teacher, if asked her pupils' average age, will probably reply by giving the most frequent age. The mode, therefore, is peculiarly dependent for its significance, and indeed for its existence, on the structure of a series. There may, indeed, be more than one mode in a series, but in educational work the presence of two or more modes indicates a probable fault in the procedure.

We may distinguish the *empirical mode* and the *theoretical mode*. The former is obtained by inspection. It is taken as the mid-value of the class-interval containing the largest number of measures. It, therefore, depends on the size and position of the classes in the frequency table. This suggests that by varying the size and position of classes we may obtain different values of the mode and by combining them, arrive at a more accurate value of it than is possible with a single determination. This method is described in most of the textbooks.[5]

The theoretical mode is not obtained directly from the data, but is the mode which would result from an indefinitely large number of measurements sub-divided into very minute classes. Its most satisfactory determination is based upon fitting an ideal frequency curve to the actual series. The value of the variable corresponding to the maximum ordinate of the fitted curve is the theoretical mode. Its determination involves some rather advanced work in mathematics, and it is not being used in educational measurements. A substitute for this method has been suggested and is said to work well. According to it

$$\text{Mode} = \text{Average} - 3\,(\text{Average} - \text{Median})$$

The average for the series given in Table I is 23.91. The median, as shown above, is 24.57. The formula just cited yields 25.89 as the value of the mode.

When the performance of pupils is recorded in the form of rates, *e. g.*, of words read per second or letters written per minute, a measure of central tendency sometimes used is not the average of the rates, but their *harmonic mean*. This may be defined as the

[5]See Bowley, Arthur L., *The Elements of Statistics*, London, 1907, pp. 118ff.; also King, Wilford I., *The Elements of Statistical Methods*, New York, 1912, pp. 122-125.

reciprocal of the average of the reciprocals of the recorded measures. Suppose five pupils read at the following rates per minute: *A*, 80 words; *B*, 100 words; *C*, 120 words; *D*, 140 words; and *E*, 160 words.[6] The average of these rates is 120 words per minute. The reciprocal of each rate expresses the *actual time* required by each pupil to read one word. Thus, *A* required 1/80 min., or 0.0125 min., to read one word; *B's* time was 0.01 min.; *C's*, 0.00833 min.; *D's*, 0.00714 min.; and *E's*, 0.00625 min. The average of these is 0.00884 min. It shows the average time required to read one word. The reciprocal of this average of reciprocals is 1/0.00884, or 113—the average number of words read per minute as computed by finding the harmonic mean. Note that we have found the average of the reciprocals of the rates and taken the reciprocal of the result. The average rate by this method is 113 words, while the average by using the rates directly is 120 words. The harmonic mean is always less than the Average. It is, therefore, clear, that results as found by different investigators are only comparable when computed by the same method.

On practical grounds it is difficult to see why we need the harmonic mean as an alternative method in computing the central tendency of rates. There is no essential difference between rates per unit of time and certain economic measures such as wages per day. It is certainly not customary to express the central tendency of the wages of a group of workmen by using the harmonic mean. In the judgment of the writer, the introduction of this method in educational reporting serves no useful purpose.

A measure of C. T., although indispensable, by no means sufficiently represents a series. It gives but one of the two chief characteristics of it. The second of these is the closeness with which the measures group about the C. T. The measure of this characteristic is called the *measure of variability,* or dispersion. There are three such measures in common use, and one of them should be used and reported, not only for its own sake, but also as a criticism of the measure of central tendency. All measures of variability, like those of C. T., are expressed in units of the series. For example, if the units of the series are words spelled, the C. T., as well as the varia-

[6] Gray, William S., *Studies of Elementary-School Reading Through Standardized Tests.* p. 15.

bility, will be expressed as a certain number of words. The *average deviation* (A. D.)—also the mean deviation, or mean variation—is simply the average of the amount by which each of the measures differs from the average, median, or mode. It is being suggested by several writers that on theoretical grounds the A. D. should be computed from the median, because when so computed for any series, it is at a minimum. People in general do not sense the meaning of the A. D. or any other measure of variability as readily as they do a measure of central tendency. The A. D. may be thought of as the amount by which every measure of the series might differ from the C. T. without influencing it.

The *standard deviation* (S. D. or σ) is another measure of variability. It is found by squaring the difference between the individual measures and the C. T., adding these squares, dividing by the number of measures, and extracting the square root of the quotient. The S. D. is theoretically the best measure of variability—at least when the series is 'normal' or nearly so. Its meaning, however, is not apparent to the lay reader, and it is difficult to explain in terms of the series. If, in Figure 2, perpendiculars are drawn from the two points on either side of the maximum ordinate at which the curve changes from concave downward to concave upward, the distance along the base line from the foot of the maximum ordinate to the foot of either perpendicular represents the S. D. Between these perpendiculars a little more than two-thirds of the area of the curve is included. In series, therefore, which approximate the normal type, we may expect about two-thirds of the measures to differ from the C. T. by not more than the standard deviation. The S. D. is at a minimum when computed from the average; and according to the best practice, it is therefore taken from that measure rather than from the median or mode.

The *quartile deviation* (Q) is half the range within which the middle half of the measures lie. In its computation one finds the three points in the range which divide the number of measures into four equal parts. These points are found in exactly the same way as the median. The points on the range at which these divisions fall are called the quartiles. The first, or lower one, is called the lower quartile, or 25-percentile (Q_1); the third or highest one is called the upper quartile, or 75-percentile (Q_3). The second one,

of course, divides the measures into the two equal parts and is the median; it is not, therefore, referred to as a quartile. The difference between the upper and lower quartile is the range within which half the measures lie. This difference divided by 2 is Q. In the case of normal distributions, it is called the *probable error* (P. E.).

It is important that the student of educational statistics should understand not only that A. D., S. D., and Q are measures of variability, but also that they are themselves important units of amount, and that for some purposes they replace the units in which the measures were recorded. Suppose that a group of eighth-grade pupils is tested in arithmetic and in handwriting. If a certain pupil scores 17 examples correct in arithmetic, and Quality 75 in handwriting, according to the Ayres scale, it is impossible to say how much the one performance is better than the other, unless both performances are expressed as a certain number of A. D's, S. D's, or Q's above or below the median of the group. Either of these units will serve to express the measures of a series in such a way as to permit comparison with the measures of another series expressed in the same units of variability, even though the original units of the two series are different.

If it is desired to compare the variabilities of two different series, we can hardly do so by using either of the three measures described above. If, for example, a group of children is rated in composition by the Nassau County Supplement to the Hillegas Scale, and in spelling by means of a list of 100 words, since the possible range of achievement in composition is only from 0 to 9, while in spelling it is from 0 to 100, the variability in the latter case will appear to be much greater than in the former. Under such circumstances, it is customary to divide the measure of variability by the C. T., the result being the so-called *coefficient of variation.* Generally speaking, it is an error in method to compare the measures of variability of two series unless the series are expressed in the same units and have approximately the same C. T.

Although the analysis of series may be carried to higher degrees of refinement, educational workers report few measures other than those of central tendency and variability. No small space, however, in educational literature is devoted to the determination of the degree of *correlation,* or mutual implication, existing between

paired measures of different traits or performances. Thus, if 100 high-school children are rated in Latin and in algebra, the extent to which an individual having high, medium, or low ratings in one subject tends also to have high, medium, or low ratings in the other, indicates the correlation for the group in question between ratings in Latin and ratings in algebra. Various methods have been devised for giving numerical expression to this correlation between two series of paired measures. Two of the methods suggested by Spearman—the "Rank Difference" and the "Footrule"—depend on the extent to which the individuals ranking first, second, third, etc., in one series tend also to rank in the same order in the other. Such a correspondence in ranking would indicate perfect correlation, and would be expressed by the integer 1. By another method (suggested by Sheppard) each measure in the two series is given a sign plus or minus, according as it is greater than or less than the central tendency, and the number of times a plus or a minus sign in one series goes with the same sign in the other series for the same individual is noted. The greater the proportion of "like-signed pairs," the greater is the correlation. The formula which is used in connection with this method again yields "1" as the measure of perfect correlation, *i. e.*, as the measure obtained when all the pairs of signs are alike.

Still other methods are used, but the so-called product-moment method is the one most commonly accepted. The meaning of the formula used in applying this method cannot be made clear in a brief description. The reader is referred to any textbook on statistics for a treatment of it. The measure of correlation yielded by this method is called the *correlation coefficient*, and its symbol is "*r*." To such an extent is the product-moment method, the standard in statistical work that coefficients derived by other methods are usually converted into it.

In determining the degree of correlation which subsists in a given group with respect to two measurable characteristics, it must be emphasized that the individual measures, pair by pair, are to be used. It is not correct procedure to break the group up into smaller sub-groups and to compare the averages in the two characteristics for each group. Averages obscure individual variations

and it is precisely the individual variations which are important in securing a correlation coefficient.

The correlation coefficient as an expression of relationship has been uncritically adopted. It is a measure of mutual implication. It fails to indicate the extent to which either of the measured characteristics depends upon the other. It may, however, be used in finding other expressions which do so. Thus, statisticians, while becoming more critical of the correlation coefficient, are using it to determine the two so-called *regression coefficients,* which permit a statement of the change in one characteristic likely to accompany a unit change in the other. I found, for example, that the correlation coefficient between achievement in answering thought and memory questions in history was +0.40. This was not nearly so informing as were the regression coefficients, one of which showed that success in answering thought questions accompanied success in answering memory questions to the extent of nearly 0.90—"1" being the maximum. The other regression coefficient, *i. e.,* of 'memory' on 'thought,' was less than 0.20.

Most of the foregoing statements, and many others not made here, depend to a large degree for their validity upon the existence of 'normal' distributions. Although the ability of school children in a given grade is presumably close to normal, measurements obtained from test material often fail to support the presumption. Aside from the fact that the pupils 'measured' are often relatively few and poorly selected, and that results are frequently scored by interested or indifferent persons, there is another important reason for this—a reason which has to do with the test itself as a measuring device.

It is necessary, in the first place, to understand clearly what it is that we are measuring when we test school children. Mr. Courtis has pointed out a useful distinction between capacity, ability, and performance.[7] *Capacity,* as he defines it, represents the possibility of development—the natural endowment of the individual. It is his potential ability, and is independent of training. The *ability* of an individual is defined as the power actually developed by the effect of training upon inherent capacity. *Performance* is the spe-

[7]Courtis, S. A., *Third, Fourth, and Fifth Annual Accountings, 1913-1916.*

cific achievement of an individual resulting from his ability as evinced under the conditions of the test.

Although we may have good reason for supposing that the distribution of school ability is closely similar to the normal distribution, we see in the distinction which has just been made, that what we are really measuring is not ability, but *performance,* and that while ability may be constant at any particular time, performance will vary according to the conditions under which the ability manifests itself. This constitutes a real difficulty in educational measurements, especially when conclusions are drawn concerning individuals rather than groups of individuals. It has been proposed that as many as twenty-five tests of a single individual ought to be made and the average of the results used before we can be reasonably certain that we have a reliable measure of performance. This is an extreme point of view and seems hardly to be tenable. Courtis has shown that if children are given a second test, only about 20 percent of them will show marked differences in performance. He has also shown that giving additional tests does not materially change the score for a class as a whole.

Since, however, what we are measuring is performance rather than ability, our tests must be so constructed that performance will tend to vary with ability, *i. e.,* to be an index of it. Thus, the conformity of the resulting measures to a normal distribution furnishes a criterion for judgment as to the adequacy of the test material. Assuming that ability is distributed approximately normally, it is highly desirable that the test material should be such as to bring out that form of distribution. Too often this is not the case. On this ground a great deal of test material now being used in education is faulty. For example, most of the testing which has been done in spelling by using Dr. Ayres' list has resulted in a large number of perfect papers, or papers that were nearly perfect. This is because the examiner has chosen words which Dr. Ayres found possessed little difficulty for the grade in question. Recently, results were published from the use of words, each of which, according to Dr. Ayres' list, had been spelled correctly by 88 percent of the children. Naturally, a very great number of the children spelled all the words correctly. Under such circumstances, no measure of central tendency could be regarded as significant. In the

fourth grade, 15 percent of the children wrote perfect papers. More of them spelled all the words right than spelled any other number of them. The test material was too easy to register variation in performance among the more capable children and therefore failed to afford any index of their ability. It is certain that among the children who spelled all the words correctly there was a wide difference in ability which would have caused a correspondingly wide difference in performance, if the material had been capable of showing it. The same inadequacy of test material is revealed when it is too difficult for the pupils to whom it is presented. Under such circumstances a large number of children of varying degrees of ability will be unable to register any performance in terms of the test. For example, in a certain test in arithmetic 145 children out of 943 were reported as unable to solve a single problem. Among these children large differences in performance corresponding to large differences in ability would have been indicated by a different test, *i. e.,* by one easy enough to enable them to accomplish something.

The best test material will be found to be midway between that which, through being too easy, fails to register variation among the pupils of greatest ability, and that which, through being too difficult, fails to register variation among those of least ability. A spelling test, for example, composed of words, each of which may be expected to be spelled correctly by 50 percent of the children in the grade taking the test, will afford ideal results. In general, test material is most adequate which yields a series of measures with an approximately normal distribution. Any piling up of measures at either the high end or the low end of the range is a defect. Such material is impossible when given to pupils in several successive grades. In the writer's judgment, no test is satisfactory when used in grades more than two years apart.

Various types of tests are being used to measure educational products in terms of the performance of children. Dr. Ayres has distinguished three. First, there are those in which the material is arranged by steps of increasing difficulty from very easy to very hard, such as Gray's Oral Reading Scale, and Woody's Arithmetic Scale. These may be called *difficulty tests.* They measure "how hard." Second, we have *accuracy tests,* or those which measure "how well." Examples of these tests are the various reading tests

which measure the quality of reading by the accuracy of the reproduction of it. Third, we have tests in which the material is intended to be of the same difficulty throughout, and in which the pupil is required to do as much work as he can in a given time. This kind of test may properly be called a *speed* or *rate test.* It measures "how much." Doubtless we should measure performance by all three of these devices. Ultimately we shall have *at each point of difficulty* as set up in the first type of tests, a large number of elements—words, problems, passages to read—each of the same difficulty. The arithmetic tests used in the Cleveland Survey are an approach to this. The criticism of the difficulty tests which have a single element at each step is that the difficulties of the elements are of different kinds. The criticism of the accuracy and rate tests is the varying difficulty of their elements. If, however, we have a large number of elements at each level of difficulty, we may then apply the kind of measurement used in the second and third types of tests, and, *within the difficulty in question,* we may measure both "how well" and "how much." Under such circumstances, our record of performance would, so to speak, consist of three dimensions. It would also afford a better index of *ability.*

Meanwhile, many new tests are needed, as well as much refinement of the material of tests now in use. It is unnecessary at this time to decide the relative merits of difficulty, accuracy, and rate tests. At present, the best criterion to apply to these types of tests is that of the regularity of the series of measures derived from them. If we are warranted, as it seems we are, in inferring that ability is distributed with an approximation to normality, then the highest requirement on the part of a test is that it measure performance in such a way as to be a true index of ability. The conditions under which tests are administered and the methods by which they are rated must be rigorously controlled, but the first requirement is that the means of measurement should be sensitive instruments, capable of registering variation in the things which they presume to measure. We may labor ingeniously at our analyses of results and may bring from afar the most potent methods which statistical theory has evolved, but we shall accomplish little if our instruments are as grossly defective as some of those which are now being employed appear to be.

CHAPTER X

TRAINING COURSES IN EDUCATIONAL MEASUREMENT

S. A. COURTIS
Supervisor of Educational Research, Detroit Public Schools

The rapid growth of the movement for measurement and the importance of the knowledge gained from surveys and experimental studies have led to a demand for definite training in measurement work. From superintendent to teachers, educational workers have been quick to recognize both the possibilities of the new tools and their own need for training. Wherever courses in measurement have been offered, the response has been surprisingly large. As a result the number of courses available in colleges and universities is steadily increasing and their influence is spreading to other fields. There are few teachers' institutes which do not touch upon some phase of measurement work, and few normal schools or other teacher training agencies which are not beginning to arrange for systematic instruction along these lines. At present there is, as might be expected, almost no agreement as to the aim or method. The courses range from the shallowest survey of the literature of the field, to the most highly technical and theoretical courses of university psychology. It may not be out of place, therefore, to give some account of the needs of teachers as seen by one who has served for several years as director of research in a large city school system.

The primary aim of all courses for teachers must be to increase the teaching power of the students. But teachers are the stuff from which principals, supervisors, and superintendents are made, so the training courses must be both broad enough, and wide and deep enough to give some knowledge of the administrative and supervisory uses of tests as well as of their instructional and diagnostic functions. For most teachers, practically the only opportunity to consider the workings of the school system *as a whole* and to think about educational problems from the supervisory and administrative standpoints is that which comes to them during the period

of training. Therefore, while the measurement work in a normal school should center around the direct application of measurement to the solution of problems of teaching, the broader aspects of the results secured must not be overlooked.

The desirable outcomes of courses in measurement, as of other instructional work, are of four different types—(1) point of view, (2) knowledge, (3) skills, (4) power. Each of these will be discussed in turn.

1. Of all the possible outcomes of training work in measurement, none is so important as the effect upon the student's *point of view*—upon his attitude, not only towards scientific experimentation in education, but also towards his educational experiences and life. For, too often, normal-school courses result merely in giving the prospective teacher the idea that everything in education is settled. A textbook is necessarily dogmatic in statement. Courses in methods and teachers in charge of training necessarily tend to stamp this as right and that as wrong. The training period is altogether too brief to do more than outline existing ideals and practices, and indicate those that are "the best." Even practice teaching under expert guidance usually serves but to give practical experience in "handling a room" and in conforming to teaching routine. Occasionally, a student of exceptional intelligence senses the contrast between ideals and existing conditions, but even his experiences seldom stir more than questionings and dull resentment.

Teaching, itself, is even more deadening to initiative. In all systems of any size the courses of study, and often the method to be used, are fixed by higher powers. The equipment, the time allowances, the size of classes, the types and abilities of the children, are all beyond the teacher's control. What is there in the training or experience of the average teacher to develop openness of mind, or give any conception of our present educational process as a crude, inefficient, wasteful makeshift, established "by guess and by gosh," and maintained by convention and social inertia?

Therefore, it should be the supreme function of measurement courses in normal training—as it has proved to be in the educational activities in the world outside the school—to give the student

the scientific attitude of mind, the critical, impersonal, and inquiring point of view. They must teach him how knowledge arises and make him feel the cost in time and labor by which the present levels of civilization have been attained. They should so clearly reveal the part that measurements and scientific methods have played in every field of human activity that he will realize their importance in education and desire to make himself proficient in their use. They must give him bases of criticism and arouse in him such a passion for truth that all his life long he will constantly seek to test, open-mindedly, disinterestedly, impersonally, the validity of all conclusions. They should lead him to regard all educational activities as problems in course of experimental solution, so that he will be ever on the watch for those significant variations which make for progress. Above all, they must so open his eyes to the wonders of the educational process, the possibilities of child development, and the relation of progress in education to progress of the world that he may have an abiding faith in the dignity and value of his profession and a burning zeal to make some contribution to the progress of the race. If the courses do this, they will be counted successful long after all technical knowledge and skill has been forgotten; if they do not, they are failures, although their graduates know every test by name and are past masters in the art of compiling tables and graphs.

2. On the side of *knowledge* there is much to be learned. The student must acquire by actual experience, knowledge of the different types of tests and the advantages and limitations of each. He must be familiar with the methods of test and scale construction and must have a first-hand experience in giving and scoring the more important of the available standard tests. He needs to know where to go for standards and comparative data, and he should have made a careful, critical study of two or three typical survey reports. He must be given, also, some experience with the variations in performance caused by changes in conditions and must learn how these are to be controlled and interpreted. More than anything else, his practical work must serve to emphasize the difference in individual children and the need of adjustment of training to such differences.

It is not enough, however, that the teacher in training have a practical working knowledge of educational tests; he must know also tests in related fields. He should be given more than a passing acquaintance with physical measurements—height, weight, lung capacity, and their significance, and simple tests of vision and hearing. He needs also some knowledge of the methods and tests employed in measuring intelligence and capacity. Further, he needs practice in the construction of rough tests and examinations, and the formulation of aims in terms of objective standards. Knowledge that is only knowledge, is vain. If the knowledge outlined above is merely memory of things read or transmitted by word of mouth, it will be of little worth. It should be knowledge derived from personal experiences.

3. The other product of experience is *skill,* and the successful course in educational measurement will have as one of its outcomes, the ability to pass certain standard tests in statistical methods; for instance, rate tests in making typical distributions, in finding averages and medians, in computing median and standard deviations, in calculating coefficients of correlations, and in drawing graphs. Even more important than these are standard tests of ability to use educational scales in a consistent manner. There should be training on some of the standardized samples which have been published in writing and composition, until a set of 20 test samples can be marked without a variation of more than half a step of the scale.

Only as the course results in a measurable proficiency in these fundamental skills should it be counted successful. Not all persons have the mental qualifications that make possible accurate judgment by means of a scale, and not all have the aptitude for statistical work; but all teachers, without exception,*need to have measured themselves against such objective standards,* both that they may know their own powers or limitations, and that they may understand the method and aims of those who have the measurement work in charge. The greatest obstacle to harmonious, cooperative work in a city school system is the misunderstanding of those who, through ignorance, misinterpret the aims of the measurement work.

4. Finally, the successful courses in measurement should result in *power* to use measurement in the solution of educational

problems. That is, for full credit, the student should be able to pass successfully three types of power tests; (1) he should have planned, measured, and compared the effects of his own and another's teaching of a specific instruction unit, like the teaching of a given twenty words in spelling, or a certain case in long division; (2) he should have devised, executed and interpreted a simple control experiment to settle some problem arising out of his practice teaching; (3) he should be able to diagnose by means of appropriate tests, and to prescribe the remedy for, the more common causes of failure in the fundamental subjects. No teacher who has had the practical experience with tests and testing that will enable him to meet these requirements will ever be willing to teach without the aid which standard tests afford.

There are five major articles in the writer's educational creed:

1. Basic experiences cannot be transmitted by instruction.
2. Understanding of the value of tools is best learned by their use as a *means to an end.*
3. Skill in the use of tools is best developed through drill.
4. Outcomes related to self-interest have greater educational potency than abstract aims.
5. All training work must be adjusted to the varying capacities and interests of individual students.

Expressed in terms of method of teaching a course in measurement these mean:

1. That the work must be "practical" in character.
2. That it must center around, and have for its purpose, the measurement, and improvement, of the practice teaching.
3. That it must consist mainly of laboratory work with only as much lecture work and reading as is necessary to connect the activities of the course with similar activities in the school system and in the educational world outside.
4. That it should consist of a series of graded exercises or projects grouped around the main topics in such a way as to provide for individual progress and differences in interest.

It seems futile to attempt to give more specifically a statement of precise topics covered, or the length of the course in terms of years or hours. Adjustment must everywhere be made to local conditions. But the material available and its importance warrants

the prophecy that in a very short time the work in measurement will be a major subject running through the entire period of training. For a six semester course (four semesters in school and two actual teaching under supervision), the following might serve as a general program. For shorter courses, the actual amount of work done would be correspondingly decreased.

Semester I. General principles of measurement, including physical and mental tests, and elements of statistical methods, with special emphasis on individual differences, the factors causing variation, and the need for control of conditions.

Semester II. Use of the simpler standard tests in measuring the effects of practice teaching.

Semester III. Measurement by means of the more complex tests, with special emphasis upon correlation of abilities and analysis of complex ability into simple elements.

Semester IV. Measurement of the results of educational experiments and of the effect of remedial work with individual children. Emphasis on the diagnostic and supervisory uses of tests, both for individuals and for a school system as a whole. Class assignments mainly individual projects, or actual participation in surveys and other practical testing work.

Semesters V and VI. Period of probationary teaching. Reports to supervisory officers to be based on the use of standard tests of the results of teaching effort.

An appropriate conclusion to this chapter is a comment on that anomaly—a course on educational measurement in which no use of measurement is made, either as a basis of adjusting the work to the abilities of individual members of the class, or as a means of measuring the efficiency of instruction. As long as educational training concerns itself with superficial conformity to conventional practices and hasty surveys of educational literature, little in the way of progress can be expected. But if the work in the training and practice schools is actually the most efficient in the system, and if there, the teacher-in-training learns to see problems as problems, and to attack them with the best tools and methods, the training which he receives will function all through his professional life. Probably upon the character and practical value of the instructions given to prospective teachers, more than upon any other one factor, depends the success of the movement for measurement and the character of its future development.

CHAPTER XI

SUGGESTIONS FOR EXPERIMENTAL WORK

GEORGE MELCHER,.
Director of Bureau of Research and Efficiency,
Public Schools, Kansas City, Missouri.

Empiricism and Science in Education

The United States has no national system of education. Each state, each municipality, each school district plans its own course of study, determines its own school organization, supervises its own school work, chooses its own plan of procedure, establishes its own standards. Our nation is called "the world's greatest experiment in democracy." With equal truth it could be called "the world's greatest experiment station in education." All kinds of educational problems are being attacked experimentally—problems in the administration, the supervision, the teaching and the financing of schools. All varieties of curricula are involved, all kinds of school equipment and appliances, and every conceivable method of teaching the various subjects.

Many communities believe that their methods of school procedure are the best. Yet they are wholly without scientific evidence in support of their belief. Thus it is often easy for an enthusiastic devotee of some new method of procedure to influence the agencies of control to such an extent that an entire school organization is revolutionized. Witness the change of a large number of elementary schools in New York City to the Gary plan of organization. We are not arguing that New York City has not improved her schools by the change; for we do not know whether she has or not. New York City does not know. No one knows. The Gary System may have many valuable contributions to make to education. The point is that the plan is new and has not been in operation long enough to have demonstrated its superiority over the conventional

school organization. Yet, in spite of that fact, the Gary system has been and is being enthusiastically copied not only in New York City but in many other communities as well. In like manner, countless other changes in school organization and school procedure are being made year after year on the basis of plausible arguments; and no one is able to demonstrate the wisdom of the change. In many cases a few years later the school returns to its original plan. Witness the more recent action at New York in abandoning the Gary plan after spending several million dollars in introducing it. Could industrial establishments succeed by such methods?

School authorities are constantly changing methods of procedure, and courses of study. This harasses teachers, annoys parents, lowers the efficiency of work, and destroys the confidence of the public in its school system. Such arbitrary orders are based merely on the personal opinions of those in authority. Even up to the present time, almost every school policy has been adopted on insufficient evidence or upon no evidence at all. In other words, personal opinion decides the destinies of school children and determines school curricula and school organization. Scientific methods have been recently introduced into industrial work, business, agriculture, and many other forms of human activity. Even the church is having its methods subjected to rigorous criticism from the standpoint of efficiency. In the management of the war, we see great nations struggling to apply every known scientific principle. Certainly education, the most important of human industries, cannot afford to neglect any opportunity to test scientifically its methods of procedure and to demonstrate the value of its results.

In this chapter I am to make a few practical suggestions that may be helpful to the teacher, principal, supervisor, or superintendent who desires to contribute his "bit" to educational progress. These suggestions will be valueless to the persons who still believe that we can measure achievements in school work by the personal opinion—frequently the offhand and prejudiced opinion—of some one individual. We know that improvement in all human activities depends very largely upon critical studies of existing conditions and upon the establishment thereby of standards to be attained at each

stage in the development of these activities. It is evident that progress in education is dependent upon a similar process. The inability of school administrators in the past to make reliable measurements has greatly retarded progress. It is true that they have been able to measure such facts as *per capita* costs, numbers and ages of graduates, the percentages of retardation, etc. Helpful as such measurements are, they fail to yield the information upon which an adequate appraisal may be based. The recent development of means of measurement, however, permits true educational experimentation.

Method of Comparison

In order to determine progress in any phase of school work, one must first analyze the situation to discover, as far as possible, the various factors that enter into it. These factors must then be studied singly and the effect of each determined. It is the analysis of these complex situations and the study of each of their factors that renders educational measurement so difficult. This is especially true in attempting to compare the efficiency of different methods of school supervision, organization, or administration.

It is probable that experimental work can render its greatest service by being applied to a study of methods of teaching and of the organization of subject matter. These problems are constantly before superintendents of schools and one or more of them can be selected each year for special study. The mere fact that they are being studied will have a beneficial effect upon the school system.

As we do not have standards in all school subjects, it is necessary to employ many other means of measuring school achievements. One device, concerning which a few suggestions will now be offered, is the 'Control Experiment.' In order that the conclusions from a control experiment may be of value, a careful plan of procedure must be followed.

First. Analyze your situation and select one factor to be studied.

Second. Select two groups of pupils approximately equal in number, in ability, and in previous training. Each group may consist of one, two, three, or more classes.

Third. At the beginning of the experiment, carefully measure the ability of each pupil in the factor under consideration.

Fourth. Select teachers who are open-minded. If possible, select those who know something of scientific methods. Especially select those who are willing to cooperate and who appreciate the value of following directions. The teachers should be as nearly equal as possible in teaching ability.

Fifth. Prepare carefully detailed instructions for all teachers who are participating in the experiment.

Sixth. Except for the one factor that is being studied, keep all the conditions in the two groups as nearly equal as possible during the progress of the experiment.

Seventh. Continue the experiment long enough for material changes to be made—several weeks, half a year, or even more may be necessary.

Eighth. At the conclusion of the experiment, carefully measure the ability of the pupils in the factor under consideration.

Ninth. Base conclusions as to relative efficiencies upon a study of gains and losses of only those pupils who were present throughout the period of the experiment.

Tenth. Allow for the effects of any varying factor other than the one under consideration.

Eleventh. Avoid conclusions from insufficient data.

Twelfth. Record and preserve the details of the procedure. It may prove to be desirable to check the conclusions by repeating the experiment.

Explanation of Steps in Plan

First. School results are complexes. Many factors usually contribute to a single result. In a control experiment only one factor should be varied and studied in order to determine the effect of that factor; as, for example, the proper distribution of time in teaching penmanship. Assume that sixty minutes a week may be devoted to the teaching of penmanship in the fifth grade. The following question arises: "Shall this time be divided into five 12-minute periods a week, four 15-minute periods, three 20-minute periods, or two 30-minute periods?" An experiment designed to

answer such a question should provide for four groups of fifth-grade pupils, each working with one of the suggested distributions of the 60 minutes per week to be devoted to penmanship. It may be that a different distribution of time would be advisable in other grades. Hence, similar questions remain to be answered for grades below and above the fifth.

A second question and one entirely different from the one above is, "What is the number of minutes a week in each grade that secures best results in the teaching of penmanship?" Note that this is an entirely different question from the first and cannot be studied at the same time without confusion of results.

Still a third problem: "To what extent can progress in writing be secured without regular drill periods in the various grades?" This problem involves entirely different elements from the first and second problems and cannot be studied in connection with them. One factor only in a situation can be successfully studied at a time.

A pupil's ability in any subject is not a single general ability but the resultant of several special abilities. For example, in arithmetic there are an almost unlimited number of abilities, as ability in adding long columns, short columns, small numbers, large numbers, etc., different abilities in subtracting, multiplying and dividing whole numbers; different abilities in operations in fractions, as in adding, subtracting, multiplying, dividing, reducing, etc. A given control experiment can deal with only a single ability or a single factor in a situation. The science of education will be perfected by solving correctly the numerous small problems involved in the educative process, as successful manufacturing is perfected by handling correctly each detail of its work.[1]

Second. In the type experiments to which I have referred in the preceding paragraph, the number of pupils is not large. Larger numbers of pupils add to the reliability of the results, although they make more difficult the control of all the factors involved and the maintenance of constant or standard conditions.

[1]The reader is referred to the January, 1912, number of the *Teachers College Record* for a detailed study of the "Separate" and "Together" methods of teaching homonyms. Here there is a splendid illustration of a well-organized and a well-conducted control experiment. The *Teachers College Record* for September, 1913, has another good example of a control experiment on the question of "Formal English Grammar as a Discipline."

Moreover, when the number of pupils is large, the handling of the statistical work becomes very laborious. Accordingly, more trustworthy results may often be secured by the repetition of the experiment in other school systems under standardized conditions.

The groups should be equal in ability because groups having low initial scores usually make large numerical gains with less expenditure of effort than do groups having high initial scores. When equal gains are expressed as percentages they seem much greater when based on low initial scores than when based on high initial scores. Percentage comparisons based upon unequal initial scores are thus very deceptive. Suppose that one fifth-grade group has an average initial speed of 50 letters a minute and the other of 75 letters a minute. An average gain of 10 letters a minute in each group appears to be an equal gain; but when expressed in percentage form it is 20 percent for the slow group and 13⅓ percent for the rapid group. Thus the slow group has the higher percentage of gain. Yet anyone at all familiar with the teaching and learning processes in school knows that a gain of 10 letters in the rapid group is a much greater achievement than is a gain of 10 letters in the slow group. Hence the importance of having the groups of equal initial ability in order to estimate rightly the improvement. It is well that the sex distribution in the groups be equal and that the previous training and home environment be as nearly equal as possible. Especially is this equality important when small groups are used. If several classes, selected at random, are used in each group, variations in conditions in the different classes will tend to neutralize each other.

Third. In measuring the skill or ability of pupils, care must be taken that the tests measure the ability that is being studied, and that they are given under uniform conditions. No detail is too trivial to be considered. For example, a principal who was giving a test in arithmetic found a room in disorder because it was temporarily in charge of a substitute who was poor in discipline. Before giving the test he administered a sharp reproof to the pupils. Already in a bad attitude, they were humiliated by the reproof and failed to respond properly to the test. An equivalent one, given a few days later in the same room, secured far better results. Not only

must directions be followed exactly, and uniformity of procedure be maintained in all rooms, but also the proper attitude of pupils must be secured. Note in Dr Briggs' study on "Formal Grammar" that he administered the test to both groups at the same time, thus insuring uniformity.

In almost every measurement of school achievement, both quantity and quality must be considered. Thus in handwriting we measure both the rate and the quality of the writing. The customary directions for a handwriting test are: "Write as well as you can at your usual rate of speed the following sentence. Write the sentence again and again until I say 'Stop'." Suppose the teacher adds as a final suggestion, "Now, do your *best*, children"; the rate will then generally be reduced materially and the quality improved only slightly. In one room this added suggestion reduced the usual speed nearly fifty percent. From these comments it is evident that the results would be more trustworthy if the same person gave both the initial and final tests. In measuring school achievements, group measurements are usually taken, since forty pupils can be tested in a group in about one-fortieth of the time required to test each pupil individually. In some cases, as in oral reading, individual tests are necessary. Usually the amount of time required to give individual tests limits such tests to a very small number of pupils.

In testing *groups* the time-limit method is generally used; that is, rate or speed is measured by the amount of work done in a given time. In such tests, absolute uniformity of time is essential. For keeping time, a stop-watch is desirable, or at least a watch with a second hand.

When *individuals* are tested, the work-limit method may be used. According to this method each pupil is given the same amount of work and his performance is measured by the time required to do it. Individual testing by the work-limit method is doubtless preferable to group testing by the time-limit method, but it is seldom practicable in ordinary classroom work.

Fourth. The preceding paragraphs indicate the wisdom of the directions with regard to teachers. Unless teachers are open-minded, they will vitiate results. A teacher who feels that she must prove

that the method or book that she is using is superior to the other method or book will unconsciously destroy uniformity of conditions. Unless teachers have a scientific attitude, they will fail to appreciate the value of many of the requirements in the experiment and are likely to feel that a failure to get highly satisfactory results is a reflection upon them rather than upon the method. To maintain their own reputations as teachers, they must make the method successful. In order to do this they prevent the standardizing of conditions. Unless teachers follow directions conclusions will be worthless.

Fifth. In addition to the detailed instructions in the hands of the teachers, conferences with them are also desirable.

Sixth. As far as possible, all factors except the one under consideration should be kept uniform or standardized for all groups. For example, in the suggested study on "Time Distribution in Teaching Handwriting," no home study or practice on handwriting outside of school should be allowed, for such practice cannot be made uniform. No attention should be given to handwriting in other subjects, otherwise varying factors will be introduced; all practice in and teaching of handwriting should be done during the sixty minutes a week. All other work should be as nearly identical as possible. In the period between the initial and final testing, the quality of the teaching of the different groups should be, as nearly as possible, the same. It is difficult to select teachers of equal ability; but by alternating them, or by repeating the entire experiment with similar alternation, the desired result may be secured. Sometimes the same teacher may instruct the several groups at different times.

Seventh. The amount of time required for a control experiment varies with the nature of the experiment. A few weeks may be sufficient to show differences in improvement in spelling or handwriting. Several months, a year, or even more, may be needed to permit a fair estimate of the value of two methods in reading, or the value of kindergarten work.

Eighth. The same care in measuring the ability of pupils at the end of the experiment must be used as was used at the beginning. The final test must be carefully chosen and must be of the

same difficulty as the initial test. The opportunity for each group to prepare for this final test should be absolutely equivalent except as to the one factor under survey.

In a general study of the school achievements in a system, it is easy to be deceived by two measures taken at different times. Suppose, for example, that you test pupils in spelling in October, using words from the Ayres' Scale. Suppose also that the teachers of your school have not had access to the Ayres' Scale prior to this time, but that after the test each teacher has a copy of it and drills on the words. Even though at the close of the experimental period you do not use the same words as in the first test, but select from the Ayres' Scale other words of equal difficulty, it is probable that your children will have apparently made great gain. You will not, however, have measured the real gain in spelling ability, since the pupils have been specially prepared for the second test.

Ninth. The scoring of results must be absolutely uniform and the tabulations made in the same way. The greatest uniformity is secured by having the same person do all the scoring and tabulating. If more than one person participates in this work, specific directions must be given to insure uniformity of work. The analysis and the correct interpretation of the gains and losses in a control experiment are the most important parts of the work. The value of a method or of a given material is measured by the gain which results from its use. The method of computing this gain will depend somewhat upon the character of the factor that is being studied. In general, in measuring school achievements, the median, as a group measure, has the advantage over other measures. It is easily and quickly computed and is not unduly affected by extreme scores. These extreme scores are always under suspicion. Furthermore, especially high or especially low individual scores have little scientific value in determining, let us say, the efficiency of a method of teaching, since what is done by the unusual pupil does not measure the value of the method for the great mass of pupils. Very high scores or very low scores affect the average much more than they do the median. The median, moreover, is likely to represent more closely the central tendency. It is often desirable to compute from the medians the percentage of gain or loss. When, however, the

percentage of gain or loss *for each pupil* is reported, the figures are misleading, unless the bases (*i. e.*, the initial scores) on which the percentages are computed are equal. Besides medians and percentages, the distribution of the scores (the frequency of each score) must receive careful consideration. The amount and percentage of gain or loss for all pupils having the same initial score are valid figures. The measure of the variability of the gains is likewise important.[2]

Tenth. After every reasonable effort has been made to keep the factors uniform during the experiment, varying factors will often enter. These must always have consideration and their probable effects must be estimated.

Eleventh. A single experiment, even when carefully conducted, is often not conclusive. In the report on the 'separate' and 'together' methods of teaching homonyms, to which reference was made above, note how guarded Dr. Pearson is in his statements. So far as his experiment is concerned, certain things are true. Another experiment might show different results.

Twelfth. When any conclusion of educational value has been reached by a single experiment, it should always be possible to verify the result by repeating the experiment at another time or in another school under the *same standardized* conditions. This can be done only when a careful record has been kept of the plan of procedure and of the important controlling factors in the experiment.

Uses of Control Experiments

Some of the uses of control experiments have been suggested in the preceding discussion. While these uses are various, the following may be particularly mentioned:

a. To determine the relative value of two methods of teaching.

b. To determine the relative value of two books or two kinds of drill material.

c. To determine the best distribution of a given teaching time.

[2]For statistical methods see Thorndike's *Introduction to Mental and Social Measurements*, Science Press, New York; Whipple's *Manual of Mental and Physical Tests*, Warwick and York, Baltimore, and Rugg's *Statistical Methods Applied to Education*, Houghton and Mifflin and Company.

d. To determine the amount of time needed to secure optimal results in a given subject.

e. To determine the relative value of different organizations of subject matter.

f. To determine the size of classes that will secure optimal results.

g. To compare two types of school organization.

Possibly these last two are too complicated to handle as an ordinary control experiment, but the principles of a control experiment should be applied.

A control experiment is simple, but requires a scientific attitude and the constant use of good judgment and common sense. Many experiments of great promise have been rendered valueless by the neglect of one or two minor details. It will be a great boon to education when the various experiments that are being made in its field are placed under sufficiently controlled conditions to render the conclusions of scientific value.

Large Problems

In addition to the simpler problems suggested above, there are many others which are too comprehensive for any school district or even for any state to solve alone, and which must, therefore, be solved, if at all, by the cooperation of many agencies. Investigations concerning types of school organization, length of sessions, sizes of classes, evening schools, recreation in and outside of school, vocational education and guidance, training of teachers, etc., are of too general a character to be satisfactorily made by a single agency. Various organizations are working on certain of these problems. Among them are the following:

A. The National Society for the Study of Education.

B. The National Education Association.

C. The National Association of Directors of Educational Research.

D. The American Association for the Advancement of Science.

E. Various national organizations for teachers and workers in a number of fields.

F. Departments of education in universities and normal schools. (Some of these institutions have organized cooperative bureaus as described in another chapter of this Yearbook.)

G. State departments, some of which have cooperative bureaus similar to those in universities and normal schools.

H. Educational Foundations.

I. United States Bureau of Education.

The United States Bureau of Education should have a Division of Educational Standards and Measurements. This bureau should be in charge of a thoroughly trained scientific student of education, seasoned by practical and successful school experience, and should become a national clearing house for educational experiment. Several lines of work are open to such a bureau.

First. Testing of Methods. Two methods of teaching could be tried in a score or more of rooms in each of one hundred cities. The United States Division of Standards and Measurements could prepare outlines of experiments, provide both the initial and final tests, compile and analyze data from these one hundred cities and deduce conclusions that would have general value in educational work. The United States Bureau of Education would thus develop an influence in the field of education similar to that exerted by the Department of Agriculture in its sphere. Bulletins of the United States Bureau of Education would then be read and studied by progressive teachers and school administrators as the Agricultural Bulletins are now read by leaders in agriculture. Such work would enable every community to profit by the experience of other communities by adopting the really successful plans and avoiding or discarding those found to be unsuccessful. Under present conditions, however, although a given method may prove unsuccessful in a score of communities this year, another group of communities may try the same method next year, not knowing that it has already failed or been discarded. Thus an unsuccessful method, if well-advertised, may gain admittance into many school systems only to be cast into the educational "junk pile" in time.

Second. Testing drill material, books, and school appliances.

Third. Conducting at various points in the United States educational experiment stations, and making preliminary tests of certain methods before extending their use to other localities. If the

vertical system of handwriting had been tested out for a few years in a large commercial center, its shortcomings would have been discovered, and it probably would not have spread over the country.

Fourth. Investigating many of the large problems of school administration that can be settled only by nation-wide study. These studies should be impartial and should not be conducted by enthusiastic advocates of the scheme to be studied. This caution is submitted because even the United States Bureau is thought to have sometimes fallen under the influence of emotionalists and educational promoters. The value of the Gary system, as referred to above, is a problem of national importance. The same Bureau of Education should be able to make a survey of the Gary Schools in operation in various places, and to secure accurate data as to the results that are being obtained. In the course of ten or twelve years, the country could know the facts with regard to this organization. Kansas City and several southern cities have a seven-year course of study in the elementary schools. If it is possible to cover the course satisfactorily in seven instead of eight years, the entire educational world should be made aware of this fact by an impartial study of the work of these seven-year systems furnished by the United States Bureau of Education. The movement to establish junior high schools is well under way. The value of this type of organization should be established, and no agency could do so better than the United States Bureau of Education.

While, therefore, there are many problems, both large and small which await solution, the value of the solution of many of them, it is urged, would be greatly enhanced if it were made from an authoritative, impersonal, and national point of view.

CHAPTER XII

A LOOK FORWARD

CHARLES H. JUDD
Director, School of Education, University of Chicago.

A paper dealing with the future can justify itself in a scientific volume of this kind only when it bases itself on an analysis of present conditions and aims to develop as a result of such an analysis suggestions for improvements and enlargements of the movement under discussion. The reader is warned, therefore, at the beginning that this paper looks backward as well as forward, in order that justification may be furnished for some of the plans urged as desirable for the future.

One fact which is evident to every student of school problems is that the movement toward the development of measurements is both promoted and seriously encumbered by a vague popular demand. Parents have heard that there are methods of finding out whether their children can spell or add or read satisfactorily, and immediately a clamor arises for a measurement of the local school. The demand is likely to be especially keen if there is some parent in the community who does not like the superintendent or the principal. Such a parent never for a moment believes that responsibility for unsatisfactory school results is to be traced to the native limitations in the ability of his child or to the home atmosphere in which the child grows up. Such a parent is quite certain that measurement will detect at some point a lack of perfection, and then he knows that his dislike for the school officer will have the sanction of science.

It is little wonder that school superintendents have often been afraid to have their schools measured. Especially hazardous is it to have the schools measured in many respects in a single survey. The number of imperfections sure to be revealed in a general survey is appalling. One has the same dread of going to a dentist to have

his teeth examined, knowing well that one's teeth are sure to 'go to pieces' under the keen scrutiny of the expert. The case would be still worse for most of us if we were obliged to submit the imperfections of our features to the relentless analysis of an expert in physiognomy.

In the presence of a popular demand for the revelation of imperfections and the absolute certainty that imperfections exist, it is not difficult to understand why there should be a tendency on the part of many school officers to combat the movement toward wide-spread measurement.

The future is sure to develop a new and more wholesome attitude on the part both of the public and of school officers. Indeed, the measurements which have been made up to this time have more than justified their cost in effort and money, because they have dispelled forever the idea that schools should produce a uniform product or one that is perfect in its attainments. We all understand now in definite scientific terms that children are different from one another, that the lower grades progress slowly toward satisfactory results, that movement within the school is purchased at great expenditure of labor on the part of all concerned, and that the best we can hope for is improvement—not absolute achievement of ideals.

With the theoretical ideal of perfection overthrown, there is now an opportunity to set up rational demands. We can venture to tell parents with assurance that their children in the fifth grade are as good as the average if they misspell fifty percent of a certain list of words. We know this just as well as we know that a certain automobile engine cannot draw a ton of weight up a certain hill. No one has a right to make an unscientific demand of the automobile or of the school.

As soon as school officers recognize the fact that measurements define for them just how much may reasonably be demanded, they will be unafraid of measurements. Indeed, they will learn the administrative lesson that it is better to know for purposes of ordinary routine what ought to be demanded than merely to guess at conditions. The writer once heard a business man put the matter very clearly. He was looking at some diagrams that showed the results

of a study of schools. "This," he said, "is the sort of thing business has learned to do. We used to be offended if anyone criticised our methods or commented on our results. Now we know that our best friend is the man who comes and tells us exactly where we stand. The one thing business cannot approve is ignorance about results. We do not fool ourselves any more, come what will of the revelation."

The school principal who knows in advance where his school is weak and where it is strong, is armed against criticism. But more than that, he is guided in his future efforts. The purely negative result that adverse judgment causes no shock, is of some importance, but the positive result that the school is stimulated to improve itself is a matter of supreme advantage. If we can devise methods of knowing ourselves, we shall take up the tasks of self-improvement with assurance and with discrimination.

The first prophecy, then, which one can venture with a good deal of assurance is that school officers will learn to anticipate popular demands and will thereby come into possession of information which will guide them in their own work.

A second general fact about measurement is that up to this time it has dealt with very broad problems and usually has grouped together great masses of results. This is seen in the fact that one speaks in a large way about medians of thousands of cases. The sheer breadth of our studies has intimidated teachers. They feel that the machinery is set up to deal with systems of schools, but not with their detailed problems.

It is natural enough that the beginnings of this science should concern themselves with broad, remote facts. So it has always been. The race developed astronomy first because celestial facts are remote and on a vast scale. It is only in the latter days of refined scientific study that we have come to know details about our own bodies and the facts of social organization.

Thanks to the energy which has been expended in scientific work, we have the gross methods well in hand. Refinement of methods has begun. Formerly we used to compare school system with school system. We shall continue this, but we can now begin to use our methods for the more specific study of individual cases.

Take, for example, the refinements which Ayres has added to his last writing scale. Along the bottom are the figures which tell the teacher in detail how speed and quality are distributed in the normal class. There is in this refinement a larger recognition of the teacher and of the classroom contact between teacher and pupil. One can not help recognizing that measurement is becoming surer of itself and is taking up details. The broad first facts have been collected and formulated. Now there is a penetration to the deeper problem —one is almost tempted to say to the real problem.

As soon as teachers learn the possibility of using definite measurement to solve their individual problems, they will share with superintendents the attitude which was described above of wanting to uncover the exact facts. Here, for example, is a difficult pupil. How far is he behind the class at the opening of the school year? How rapidly does he progress? Whatever the answer, the teacher will be aided in directing the pupil's work if that answer can be known with definiteness and detail.

Everywhere there are indications that measurements are to be used by the teacher. The results of supervised study are being measured. The results of different methods of teaching are being accurately determined. Thus, different systems of reading, different methods of teaching long division, and different methods of manipulating the decimal point are being studied.

Up to this time, teachers, partly because they shared the dread of measurements and partly because they thought of measurements as remote, have stood aloof from the movement. Now there appear the beginnings of a tendency to make measurement a part of the class routine. The arithmetic lesson serves at once as a drill exercise and as an opportunity for measuring results. The rhythmical beating of time in writing helps in the formation of a habit and tells the teacher what members of the class, if any, are lacking in skill. The measurement of rate in reading helps the teacher to decide which members of the class require special attention.

These beginnings mark the path along which the measurement movement must travel in the coming years. There is need of new energy in devising methods of class routine which will bring to the teacher the exact results which will show how successful has been

the work of the class and of the individual. Those who complain that the teacher does not have time to make measurements miss the point entirely. The teacher often wastes time and effort under existing conditions because of ignorance of the direct point where application of teaching energy would be most effective. The right kind of classroom measurements, as suggested in the examples cited above, do not interrupt class routine at all, but contribute exact methods of procedure at the same time that they reveal to the teacher where the class stands.

A conception such as that given in the last paragraphs will also clear up another difficulty which teachers sometimes point out. They complain that the volume of experimentation is so great that the class exercises are disorganized and disrupted. The advice which ought to be given to a teacher who makes this complaint is that one kind of class exercise should be transformed at a time. Methods should be built up in each subject which serve both the purposes of measurement and of teaching. This can be done, but it requires readjustment and planning.

The second prophecy which one may venture is, accordingly, that measurement will more and more take up details and will become a common instrument in the hands of the classroom teacher.

One objection which has been urged again and again against measurement is that it deals only with the formal and mechanical aspects of education. This objection has nowhere been more definitely stated than by Superintendent Horn in his *Supplementary Survey of Portland Public Schools* where he writes as follows:

"It should furthermore be kept in mind that there are many things about a school system which can never be definitely measured or stated with mathematical accuracy. Just where the line is to be drawn between the measurable and the non-measurable elements that enter into a school is a matter concerning which there is much difference of opinion. In other words, the element of opinion enters to some extent even into the matter of the possibility of measurement.

"For instance, it is an undoubted fact that any man can go into a city and count the school houses or the number of the desks. Any man can find out the number of teachers employed. Any man can count for himself the number of pupils present in a given room.

"It takes no particular ability to enable an inquirer to find out just how much money is being spent. If the schools spent nine hundred thousand dollars in one year and a million dollars the next year, any one can deduce the fact that they spent one hundred thousand dollars more the second year than the first year.

"On the other hand, after a comparatively few such facts have been definitely ascertained, we come to subjects that cannot be measured in mathematical terms, and concerning which there are no definite standards. In this realm ideals are not always definitely established and opinions are almost certain to vary widely.

"For instance, if you take two classes, one across the hall from the other, who can decide in which of the classes the higher ideal of truthfulness or honesty prevails? Who can say which teacher is more successful in making the children self-reliant, and to what extent? We all know that in such a case, if the two teachers are both fairly good, many pupils and patrons will consider one the better teacher, while many others will consider the other the better. Especially will this be true with reference to such matters as the teaching of honesty, industry and self-reliance. Incidentally, these very things are recognized as being among the most important of all the elements entering into the question of the teacher's efficiency. A school that turns out manly, honorable, self-reliant boys and womanly, efficient girls is likely to be at least a fairly good school, no matter what it may do otherwise. A school that fails to turn out such pupils can hardly be considered a good one, no matter what it may do for its pupils in the way of reading, or writing, or arithmetic. And yet these very things, which may decide between the success or failure of the school, are matters which it is almost impossible to estimate accurately, and concerning which there may be a wide amount of honest difference of opinion."[1]

The success of the measurement movement depends on its ability to meet this type of objection.

Some of us might be entirely willing to rest the case after asking whether in practical school life anyone ever saw a teacher thoroughly competent in teaching ideals but neglectful of reading and arithmetic. The fact is that the conscientious teacher always gives attention to both, and the successful teacher is able without omitting one to cultivate the other. The theoretical possibility of thinking of the two results separately has little significance in dealing with real teachers and real schools. Good reading is a school virtue, and

[1] P. W. Horn, *Report of Supplementary Survey of Portland Public Schools*, pp. 6-7. April, 1917.

when one has measured good reading, he has measured more than the trivial or formal side of education.

The hope of the measurement movement is, however, to do more than to deny the validity of such criticism as Mr. Horn makes. There is to be progress in covering more fully the details of school work. Today we know how to measure many aspects of teaching. The reason for our early attack on the formal elements is that these yield readily to analysis and thus to theoretical isolation and exact treatment. What we need to do is to carry our analysis further and then new measurements will become easier.

A few years ago reading tests seemed impossible. Today we have mastered the distinction between oral and silent reading. We have good methods of measuring some of the more common types of deficiency and we know the rate of progress which is normal in the more obvious phases of interpretation. The progress in this field within a single year is so large that there is nothing but optimism in the minds of those carrying on the work. What we need is more interest on the part of practical workers and more experimentation with methods.

Those of us who have watched the progress of measurement will recall distinctly that the earliest critics of the movement were more emphatic than the present-day critics in declaring that school results could not be measured. This type of criticism was the one with which Mr. Rice's opponents thought they had forever eliminated him and his type of work from the schools. Steadily the range of measurements has broadened. Steadily the productivity of the movement has increased. It is not for the advocate of the movement to prophesy its limits; it is perfectly safe for him, however, to assure all the world that the end is not yet in sight. So long as advantage comes from the pushing forward of this movement, so long as ingenuity is at hand to devise new modes of procedure, the answer to the objection that measurement is limited to a few trivial aspects of teaching is steadily becoming more cogent.

This hopeful conclusion is fully supported by one fact which serves at the same time to reveal one of the most important advantages of measurement, namely, the fact that with the develop-

ment of measurement there is coming into education a greater general clearness and definiteness of purpose.

An example will show what is meant. In the high school of Kansas City, Kans., there is a system of telling the students definitely and in detail what they must do if they want to secure the higher grades, *A* or *B*, in a given course. The very fact that students in that school have all along been given *A*, *B*, and *C* shows that measuring of all sorts of intellectual and moral qualities has been going on. The interesting fact is that in most places the measuring is vague and often unsatisfactory, because no one has taken the pains to define what is wanted. Students know that teachers are often arbitrary, and, be it confessed, teachers also know that they are vague. The system referred to above removes some of the ambiguities. It improves the measuring system, making it definite and exact, because it analyzes and defines the elements of work demanded.

Suppose that the teachers of a school should concentrate for half a year on cultivating the power of concentration of attention. Is there any doubt that much new information would be gained about concentration and that there would be more accurate methods of determining its degree? Measurement will be extended in the future. The reasons why one can be so sure about this statement are to be found in the history of the past few years.

The third prophecy that can be made is, therefore, that the scope of measurement will be widened until it is sufficiently inclusive to satisfy even the most exacting critic. Concentration of attention, ability to attack various kinds of problems, clearness of insight, power of inference in various fields will be measured. The demand again is for workers who will give themselves the training in analysis and take the pains in collecting material that is necessary to bring about this consummation.

It may be unscientific to prophesy about the remoter social consequences of such a movement as we are discussing, but certain final observations may serve to show why the advocates of measurement in education are unlimitedly optimistic. The time is rapidly passing when the reformer can praise his new devices and offer as the reason for his satisfaction, his personal observation of what was

accomplished. The superintendent who reports to his board on the basis of mere opinion is rapidly becoming a relic of an earlier and unscientific age. There are indications that even the principals of elementary schools are beginning to study their schools by exact methods and are basing their supervision on the results of their measurements of what teachers accomplish. A social change of this kind is adequate justification for any movement and a sufficient guarantee for its continuance.

CHAPTER XIII

EDNA BRYNER
Russell Sage Foundation, New York City

A SELECTED BIBLIOGRAPHY OF CERTAIN PHASES OF EDUCATIONAL MEASUREMENT

These magazine articles, bulletins, reports, books and surveys are grouped on the basis of similarity of general content and are arranged alphabetically within each division according to authors.

Divisions

A. Theory of Educational Measurement and Development of the Movement.
B. Tests and Scales in Various School Subjects.
C. General Reports on the Use of Tests and Scales in Schools
D. Lists of Tests and Scales
E. Correlations between Abilities
F. Teachers' Measurement
G. Articles about Surveys and Lists of Surveys
H. City Surveys
I. State, County and other surveys.

Abbreviations

AmEd: American Education
AmJPs: American Journal of Psychology
AmScBdJ: American School Board Journal
BrJPs: British Journal of Psychology
Ed: Education
EdAdm: Educational Administration and Supervision
EdPsMon: Educational Psychology Monographs
ElScJ: Elementary School Journal
ElScTch: Elementary School Teacher
EngJ: English Journal
IndianaConfMes: Annual Conference on Educational Measurements, Indiana University
IndianaUnSd: Indiana University Studies
JEdPs: Journal of Educational Psychology

JEPd: Journal of Experimental Pedagogy
KansasBurMes: Studies by the Bureau of Educational Measurement Kansas State Normal School
NaEdAsBu: National Education Association Bulletin
PdSe: Pedagogical Seminary
ProcDepSupt: Proceedings Department Superintendence National Education Association
PopSciM: Popular Science Monthly
ProcNaEdAs: Proceedings National Education Association
PsCl: Psychological Clinic
ScHomeEd: School and Home Education
Sci: Science
ScR: School Review
ScRMon: School Review Monographs
ScSciMat: School Science and Mathematics
ScSoc: School and Society
Tch: Teaching
TchCollConEd: Teachers College Contributions to Education
TchCollRec: Teachers College Record
UnIowaExtnBu: University of Iowa Extension Bulletin
USBurEdBu: United State Bureau of Education Bulletins
YbNaSoc: Yearbook of the National Society for the Study of Education

DIVISION A

THEORY OF EDUCATIONAL MEASUREMENT AND DEVELOPMENT OF THE MOVEMENT

1. Alexander, C. Standard tests as an aid in supervision. *AmScBdJ*, 54: 1917, 17-18, 66-67, 69.
2. Ayres, L. P. Measuring educational processes through educational results. *ScR*, 20: 1912, 300-309.
3. Ayres, L. P. The economy of time through testing the course of study and time allotment. *ProcNaEdAs*, 1913, 241-246.
4. Ayres, L. P. Making education definite. *Second IndianaConfMes*, 13: 1915, 85-96.
5. Ayres, L. P. The measurement of educational processes and products. Second *IndianaConfMes*, 13: 1915, 127-133.
6. Ayres, L. P. Significant developments in educational surveying. *Proc DepSupt*, 4: May, 1916, 104-111.
7. Bagley, W. C. The need of standards for measuring progress and results. *ProcNaEdAs*, 1912, 634-639.
8. Ballou, F. W. The function of a department of educational investigation and measurement in a school system. *ScSoc*, 1: Feb. 6, 1915, 181-190.
9. Ballou, F. W. Improving instruction through educational measurement. *EdAdm*, 2: June, 1916, 354-367; also *NaEdAsBu*, 4: 1916, 196-203.
10. Ballou, F. W. Training normal-school seniors in educational measurement. *ScSoc*, 5: Jan. 20, 1917, 61-70.
11. Birch, T. B. Standard tests and scales of measurement. *PsCl*, 10: April, 1916, 49-57.
12. Bliss, D. C. School measurements and school administration. *EdAdm*, 1: Feb., 1915, 77-88.

13. Bliss, D. C. The application of standard measursements to school administration. *Fifteenth YbNaSoc,* Pt. I, 1916, 69-78.
14. Bobbitt, F. Some general principles of management applied to the problems of city school systems. *Twelfth YbNaSoc,* Pt. I, 1913, 7-96.
14a. Bowley, A. L. *Elements of statistics.* New York, 1907.
14b. Bowley, A. L. *An Elementary Manual of Statistics.* London, 1915.
15. Buckingham, B. R. Principles of scale derivation with special application to arithmetic, geography, history and grammar, with round table. *Third IndianaConfMes,* 2: Feb., 1917, 49-84.
16. Colvin, S. Marks and the marking system as an incentive to study. *Ed,* 32: May, 1912, 560-572.
17. Courtis, S. A. The comparative test as an educational ruler. *AmEd,* 15: Sept., 1911, 13-18.
18. Courtis, S. A. Measuring the child's capacity. *Normal Instructor and Primary Plans,* 1916: Oct., 31, 76; Nov. 53-54, 82; Dec. 45-46.
19. Courtis, S. A. Objective standards as a means of controlling instruction and economizing time. *ScSoc,* 1: Mar. 27, 1915, 433-436.
20. Courtis, S. A. Supervisory control by means of objective standards. *Second IndianaConfMes,* 13: Oct., 1915, 37-68.
21. Courtis, S. A. Educational diagnosis. *Second IndianaConfMes,* 13, 1915, 135-171; also *EdAdm,* 1: Feb., 1915, 89-116.
22. Courtis, S. A. Standardization of teachers' examinations. *ProcNaEdAs,* 1916: 1078-1086.
23. Cubberley, E. P. *Public School Administration.* Boston, 1916, 324-343.
24. Cubberley, E. P. Standard tests in the work of school administration: round table. *Third IndianaConfMes,* 2, 1917, 21-34.
25. Cubberley, E. P. Significance of educational measurements. *Third IndianaConfMes,* 2: 1917, 6-20.
25a. Elderton, W. P. and Ethel M. *Primer of Statistics.* London, 1912.
26. Haggerty, M. E. Some uses of educational measurements. *ScSoc,* 4: Nov. 18, 1916, 762-771.
27. Haggerty, M. E. Measurement and diagnosis as aids to supervision. *ScSoc,* 6: Sept. 8, 1917, 271-285.
28. Hall-Quest, A. L. Measuring the ultimate product. *ScSoc,* 1: May 29, 1915, 767-773.
29. Kelley, T. L. Educational guidance: an experimental study in the analysis and prediction of ability of high-school pupils. *TchCollConEd,* No. 71: 1914, pp. 116.
29a. King, W. I. The elements of statistical method. New York 1912.
30. Kirkpatrick, E. A. Measurements, standards and teaching. *ScSoc,* 1: Feb. 20, 1915, 278-281.
31. Learned, W. S. The school system as an educational laboratory. *Harvard-Newton Bulletins,* No. 1: July 1914, Cambridge, Mass. pp. 50.
32. Littell, H. V. Testing efficiency in schools. *Proceedings 28th Annual Meeting Associated Academic Principals and Council of Elementary Principals and Teachers, State of New York,* 1912, 67-75.
33. Manahan, J. L. Standard educational tests and educational administration. *High School Quarterly,* 5: July, 1917, 247-250.
34. Maxwell, W. H. Efficiency of schools and school systems. *ProcNaEdAs,* 1915, 395-402.
35. Melcher, G. The two phases of educational research and efficiency in the public schools. *ProcDepSupt,* 4: May, 1916, 183-188.
36. Meriam, J. L. Measuring school work in terms of life out of school. *ScSoc,* 5: Mar. 24, 1917, 339-342.

37. Monroe, W. S. Standard tests and their uses. *Tch,* 1: Nov. 15, 1914, 9-14.
38. Rice, J. M. *Scientific Management in Education.* New York, 1913, pp. 282.
39. Rugg, H. O. The experimental determination of mental discipline in school studies. *EdPsMon,* No. 17: 1916, pp. 132.
39a. Rugg, H. O. *Statistical methods applied to education.* New York, 1917. 410 pp.
40. Scott, F. N. Efficiency for efficiency's sake. *ScR,* 23: Jan., 1915, 34-42.
41. Spaulding, F. E. The application of the principles of scientific management. *ProcNaEdAs,* 1913, 259-279.
42. Spaulding, F. E. Measurement of efficiency in elementary and secondary schools. *Ed,* 34: Dec., 1913, 225-248.
43. Starch, D. Standard tests as aids in the classification and promotion of pupils. *Fifteenth YbNaSoc, Pt. 1,* 1916, 143-148.
44. Stockton, J. L. *Exact Measurements in Education.* Chicago and New York, 1915, 57 pp.
45. Strayer, G. D. Measuring results in education. *JEdPs,* 2: 1911, 3-10.
46. Strayer, G. D. Is scientific accuracy possible in the measurement of the efficiency in instruction? *Ed,* 34: Dec., 1913, 249-258.
47. Strayer, G. D. The use of tests and scales of measurement in the administration of schools. *ProcNaEdAs,* 1915, 579-582.
48. Strayer, G. D. and Thorndike, E. L. *Educational administration: quantitative studies.* New York, 1913, pp. 391.
49. Thompson, J. E. Educational measurements. *ScHomeEd:* Nov., 1916, 69-72.
50. Thorndike, E. L. Quantitative investigations in education. *ScRMon,* No. 1: 1911, 33-52.
51. Thorndike, E. L. The measurement of educational products. *ScR,* 20: May, 1912, 289-299.
52. Thorndike, E. L. *An Introduction to the Theory of Mental and Social Measurements.* New York City, 1913, pp. 277.
53. Thorndike, E. L. Units and scales for measuring educational products. *Proc. Conference on Educational Measurements, Indiana University Bulletin,* 12, No. 10, 1914, 128-141.
54. Thorndike, E. L. Individual differences and their causes. *Proc. Conference on Educational Measurements, Indiana University Bulletin,* 12, No. 10; 1914, 59-76.
55. Thorndike, E. L. Round table: the use of the new educational scales in ordinary school practice. *Proc. Conference on Educational Measurements, Indiana University Bulletin,* 12, No. 10; 1914, 142-154.
56. Whipple, G. M. The use of mental tests in the school. *Fifteenth YbNa Soc. Pt.I,* 1916, 149-159.
57. Woody, C. Tests and measures in the schoolroom and their value to teachers. *ScSoc,* 6: July 21, 1917, 61-66; also *Northwest Journal of Education,* 28: Dec., 1916, 154-158.
57a. Zizek, Franz. *Statistical Averages.* Translated by W. M. Persons. New York, 1913.

DIVISION B

TESTS AND SCALES IN THE VARIOUS SCHOOL SUBJECTS*

Algebra

58. Childs, H. G. The measurement of achievement in algebra. *Third IndianaConfMes,* 2: 1917, 171-183.

*For further information concerning addresses of those whose tests are listed here, see Ch. VII.

59. Childs, H. G. *Indiana Algebra Tests.* Columbia City, Ind.
60. Monroe, W. S. A test of the attainment of first-year high-school students in algebra. *ScR*,23: Mar., 1915, 159-171.
61. Monroe, W. S. Measurements of certain algebraical abilities. *ScSoc*,1: Mar. 13, 1915, 393-395.
62. Monroe, W. S. Standard research tests in algebra. *KansasBurMes.*
63. Rugg, H. O. *Standardized Tests in First-year Algebra.* University of Chicago, Chicago, Ill.
64. Rugg, H. O. The experimental determination of standards in first-year algebra. *ScR*, 24: Jan., 1916, 37-66.
65. Rugg, H. O. and Clark, J. R. *Standardized Practice Exercises in First-year Algebra.* University of Chicago, Chicago, Ill.
66. Rugg, H. O. and Clark, J. R. Standardized tests and the improvement of teaching in first-year algebra. *ScR* 25: Feb. and Mar., 1917, 113-132 and 196-213.
67. Rugg, H. O. and Clark, J. R. A co-operative investigation in the testing and experimental teaching of first-year algebra. *ScR*, 25: May, 1917, 346-349.
68. Rugg, H. O. and Clark, J. R. The improvement of ability in the use of the formal operations of algebra by means of formal practice exercises. *ScR*, 25: Oct., 1917, 546-554.
69. Stromquist, C. E. *Preliminary Algebra Tests.* Univ. of Wyoming, Laramie, Wyoming.

Arithmetic

70. Ashbaugh, E. J. The arithmetical skill of Iowa school children. *UnIowa ExtnBu*, 24: Nov., 1916, pp. 63.
71. Baldwin, B. T. The application of the Courtis tests in arithmetic to college students. *ScSoc*, 1: Apr. 17, 1915, 569-576.
72. Ballard, P. B. Norms of performance in the fundamental processes of arithmetic. *JEPd*, 2: Dec., 1914, 396-405; vol. 3, Mar., 1915, 9-20.
73. Ballou, F. W. Educational standards and educational measurements, with particular reference to standards in the four fundamentals of arithmetic. *Bulletin No. 3, School Document No. 10*, 1914. Boston, pp. 21.
74. Bell, J. C. A class experiment in arithmetic. *JEdPs*, 5: Oct., 1914, 467-470.
75. Bonser, F. G. The reasoning ability of children of the fourth, fifth and sixth school grades. *TchCollConEd*, 37: New York City, 1910, pp. 133.
76. Bonser, F. G. The selective significance of reasoning ability tests. *JEd Ps*, 7: Apr., 1916, 187-200.
77. Boston, Department of Educational Investigation and Measurement. Provisional minimum standards in addition, subtraction, multiplication and division for pupils in grades 4 to 8. *Bulletin No. 2, School Document No. 9*, 1914, Boston, pp. 37.
78. Boston, Department of Educational Investigation and Measurement. Arithmetic. Determining the achievement of pupils in addition of fractions. *Bulletin No. 7, School Document No. 3*, 1916, Boston, pp. 28.
79. Boston, Department of Educational Investigation and Measurement. Arithmetic: the Courtis standard tests in Boston, 1912-1915. An appraisal. *Bulletin No. 10, School Document No. 15*, 1916, Boston, pp. 48.
80. Brown, J. C. An investigation on the value of drill work in the fundamental operations of arithmetic. *JEdPs*, 2: Feb., 1911, 81-88; 3: Nov., 1912, 485-492; Dec., 1912, 561-570.

81. Buckingham, B. R. The Courtis tests in the schools of New York City. *JEdPs*, 5: Apr., 1914, 199-214.
82. Buckingham, B. R. Notes on the derivation of scales in school subjects: with special application to arithmetic. *Fifteenth YbNaSoc, Pt. I*, 1916, 23-40.
83. Bush, M. G. The fundamental number facts. *ScSoc*, 6: Sept. 1, 1917, 265-268.
84. Cajori, F. A new marking system and means of measuring mathematical abilities. *ScSciMat*, 14: Apr., 1914, 283-293.
85. Chapman, J. C. A study of initial spurt in the case of addition. *JEdPs*, 6: Sept., 1915, 419-426.
86. Chase, S. E. Waste in arithmetic. *TchCollRec*, 18: Sept., 1917, 360-370.
87. Childs, H. G. A half-year's progress in the achievement of one school system as measured by the Courtis tests, Series B. *Fifteenth YbNaSoc, Pt.I*, 1916, 83-90.
87a. Cobb, Margaret V. A preliminary study of the inheritance of arithmetical abilities. *JEdPs*, 8: Jan., 1917, 1-20.
88. Cole, L. W. Adding upward and downward. *JEdPs*,3: Feb., 1912, 83-94.
89. Conard, H. E. and Arps, G. F. An experimental study of economical learning. *AmJPs*, 27: Oct., 1916, 507-529.
90. Counts, G. S. Arithmetic tests and studies in the psychology of arithmetic. *Supplementary Educational Monographs*, 1, No. 4: Aug. 27, 1917, pp. 125.
91. Courtis, S. A. Measurement of growth and efficiency in arithmetic. *ElScTch*, 10: Oct. and Dec., 1909, 58-74 and 177-199; 11: Dec., 1910, and Mar., 1911, 171-185 and 360-370.
92. Courtis, S. A. Standard scores in arithmetic. *ElScJ*, 12: Nov., 1911, 127-137.
93. Courtis, S. A. The Courtis tests in arithmetic. *Final Report, Committee of School Inquiry, Board of Estimate and Apportionment, New York City*, I: 1911-1912, pt. 2, 389-546.
94. Courtis, S. A. First annual accounting of the Courtis Standard Research Tests, 1911-12; also, reprints of articles published during the year. *Bulletin No. 1, Department of Cooperative Research*, Detroit, Mich., pp. 31.
95. Courtis, S. A. Second annual accounting of the Courtis Standard Research Tests, 1912-1913; also, Standard scores and teaching of the tables. *Bulletin No. 2, Department of Cooperative Research*, Detroit, Mich., pp. 31.
96. Courtis, S. A. The reliability of single measurements with standard tests. *ElScJ*, 13: March and June, 1913, 326-345 and 486-504.
97. Courtis, S. A. Third, fourth and fifth annual accounting of the Courtis Standard Research Tests, 1913-1916; also Capacity, ability and performance in relation to standard scores and summary of tabulations. *Bulletin No. 4, Department of Cooperative Research*, Detroit, Mich., pp. 111.
98. Courtis, S. A. Courtis tests in arithmetic: value to superintendents and teachers. *Fifteenth YbNaSoc, Pt. I*, 1916, 91-106.
99. Courtis, S. A. Research work in arithmetic. *EdAdm*, 3: Feb., 1917, 61-74
100. Courtis, S. A. *Arithmetic Tests, Series A and Series B.* Detroit, Mich.
101. Courtis, S. A. *Practice Tests in Arithmetic.* Yonkers, N. Y., 1917.
102. Davies, G. R. Elements of arithmetical ability. *JEdPs.* 5: Mar., 1914, 131-140.

103. Dockerill, W. H. A. and Fennings A. J. A new test of reasoning. *JEPd*, 2: June, 1914, 356-361.
103a. Drushel, J. A. A study of the amount of arithmetic at the command of high-school graduates who have had no arithmetic in their high-school course. *ElScJ*, 17: May, 1917, 657-661.
103b. Fox and Thorndike. The relationship between the different abilities in the study of arithmetic. *TchCollConEd.* Feb., 1913, 32-40.
104. Gist, A. S. Errors in the fundamentals of arithmetic. *ScSoc*, 6: Aug. 11, 1917, 175-177.
105. Gray, P. L. Norms of performance in the fundamental processes of arithmetic. *JEPd*, 3: June, 1916, 310-318.
106. Green, J. A. Normal performances in fundamental arithmetical processes. *JEPd*, 3: June, 1916, 318-322.
107. Haggerty, M. E. Arithmetic: a cooperative study in educational measurements. *IndianaUnSd*, No. 27: *IndianaUnBu*, 12: 1915, 385-508.
108. Haggerty, M. E. (Ed.) Studies in arithmetic. *IndianaUnSd*, 3, Sept. 1916, pp. 109.
109. Hahn, H. H. and Thorndike, E. L. Some results of practice in addition under school conditions. *JEdPs*, 5: Feb., 1914, 65-84.
109a. Hanus, P. H. and Gaylord, H. D. Courtis arithmetic tests applied to employees in business houses. *EdAdm*, 3: Nov., 1917, 505-520.
109b. Heck, W. H. The efficiency of grammar-grade pupils in reasoning tests in arithmetic at different periods of the school day. *JEdPs*, 5: 1914, 92-95.
109c. Heilman, J. D. and Shultis, F. W. A study in addition. *Research Bulletin No. 1*, State Teachers College of Colorado, Greeley, 1916.
110. Hill, D. S. Measurements by the Courtis arithmetic tests. *Annual Report of the New Orleans Public Schools*, 1915-16, Pt. 3, 74-77.
111. Hinkle, E. C. Tests for efficiency in arithmetic. *Educational Bimonthly*, 8: Dec., 1913, 123-135.
112. Hinkle, E. C. Experiments for efficiency in arithmetic. The Courtis tests. *Educational Bi-monthly*, 8: Feb., 1914, 189-201.
113. Holloway, H. V. *The Relative Difficulty of the Tables.* State Gazette Pub. Co., Trenton, N. J.
113a. Howell, H. B. *A Fundamental Study in the Pedagogy of Arithmetic.* New York, 1914, pp. 328.
114. Jessup, W. A. Current practices and standards in arithmetic. *Fourteenth YbNaSoc, Pt. I*, 1915, 116-130.
114a. Jessup, W. A. Economy of time in arithmetic. *ElScTch*, 14: June, 1914, 461-576.
114b. Jessup, W. A. Grade for introduction of a text in arithmetic. *ElScJ*, 15: Nov., 1914, 162-166.
115. Jessup, W. A. and Coffman, L. D. *The Supervision of Arithmetic.* New York, 1916, pp. 225.
116. Judd, C. H. *Cleveland Arithmetic Tests.* Univ. of Chicago, Chicago, Ill.
117. Kelly, F. J. Teacher's marks: their variability and standardization. *TchCollConEd*, 66: 1914, pp. 139.
118. Kerr, Mary. Use of Courtis tests in central school building, Bloomington, Ind. *IndianaConfMes*, 12: 1914, 90-94.
119. Kirby, T. J. Practice in the case of school children. *TchCollConEd*, 58: 1913, pp. 98.
120. Kirkpatrick, E. A. An experiment in memorizing vs. incidental learning. *JEdPs*, 5: Sept., 1914, 405-413.
121. Lane, H. A. Standard tests as an aid to supervision. *ElScJ*, 15: March, 1915, 378-386.
122. McDougle, E. C. A contribution to the pedagogy of arithmetic. *PdSe*, 21: June, 1914, 161-218.

122a. Mead, C. D. *An Experiment in the Fundamentals.* Yonkers, 1917.
123. Mead, C. D. and Sears, Isabel. Additive, subtractive, and multiplicative division tested. *JEdPs,* 7: May, 1916, 261-270.
124. Monroe, W. S. An experimental and analytical study of Woody's arithmetic scales, Series B. *ScSoc,* 6: Oct. 6, 1917, 412-420.
125. Monroe, W. S. A report of the use of the Courtis Standard Research Tests in arithmetic in twenty-four cities. *KansasBurMes,* 4: No. 8, pp. 94.
125a. Monroe, W. S. The ability to place the decimal point in division. *ElScJ,* 18: Dec., 1917, 287-293.
126. Otis, A. S. and Davidson, P. E. The reliability of standard scores in adding ability. *ElScTch,* 13: Oct., 1912, 91-105.
127. Phelps, C. L. A study of errors in tests of adding ability. *ElScTch,* 14: Sept., 1913, 29-39.
128. Phillips, F. M. A comparison of the work done in successive minutes of a ten-minute practice period in the fundamentals of arithmetic. *JEdPs,* 7: May, 1916, 271-277.
128a. Phillips, F. M. Value of daily drill in arithmetic. *JEdPs,* 4: 1913, 159.
129. Rice, J. M. Educational research: arithmetic. *The Forum,* 34: Oct., 1902 to June, 1903, 281-297, 437-452, 588-607. Reprinted in *Scientific Management in Education,* New York, 1913.
130. Scott, W. Errors in arithmetic. *JEPd,* 3: June 5, 1916, 296-307.
131. Smith, A. O. Preliminary reports of the Lough-Smith tests in arithmetic. *AmED,* 17: March, 1914, 410-411.
132. Smith, J. H. Individual variations in arithmetic. *ElScJ,* 17: Nov., 1916, 195-200.
133. Starch, D. Transfer of training in arithmetical operations. *JEdPs,* 2: 1911, 306-310.
134. Starch, D. A scale for measuring ability in arithmetic. *JEdPs,* 7: April, 1916, 213-222.
135. Starch, D. *Starch's Arithmetic Test.* Madison, Wis.
136. Starch, D. and Elliott, E. C. Reliability of grading work in mathematics. *ScR,* 21: April, 1913, 254-259.
137. Stone, C. W. Problems in the scientific study of the teaching of arithmetic. *JEdPs,* 4: Jan., 1913, 1-16.
138. Stone, C. W. Arithmetical abilities and some factors determining them. *TchCollConEd,* 19: 1908, pp. 101.
139. Stone, C. W. *Standardized Reasoning Tests in Arithmetic and How to Utilize Them.* New York City, 1916, pp. 24.
140. Stone, C. W. *A Reasoning Test in Arithmetic.* New York City.
140a. Sutton, C. W. The arithmetical abilities of school children as shown by the Courtis tests. *Division of Reference and Research, Cleveland, Ohio. Bulletin No. 1,* Nov., 1917.
141. Taylor, E. H. A comparison of the arithmetical abilities of rural and city school children. *JEdPs,* 5: Oct., 1914, 461-466.
142. Thompson, T. E. *Thompson's Minimum Essentials in Arithmetic.* Manrovia, Calif.
143. Thorndike, E. L. Means of increasing efficiency in arithmetic. *IndianaUnBu,* 12: 1914, 9-25.
144. Thorndike, E. L. The relation between speed and accuracy in addition. *JEdPs,* 5: Nov., 1914, 537.
145. Thorndike, E. L. Measurements of ability to solve arithmetical problems. *PdSe,* 21: Dec., 1914, 495-503.
145a. Thorndike, E. L. Practice in the case of addition. *AmJPs,* 21: 1910, 483.
145b. Uhl, W. L. The use of standardized materials in arithmetic for diagnosing pupils' methods of work. *ElScJ,* 18: Nov., 1917, 215.

146. Winch, W. H. Some new reasoning tests suitable for the mental examination of school children. *BrJPs*, 7: Sept., 1914, 190-225.
147. Wood, E. R. Tests in efficiency in arithmetic. *ElScJ*, 17: Feb., 1917, 446-453.
148. Woody, C. Measurements of some achievements in arithmetic. *ScSoc*, 4: Aug. 19, 1916, 229-303.
149. Woody, C. *Measurements of Some Achievements in Arithmetic.* New York City, pp. 63.
150. Woody, C. *Arithmetic Scales.* New York City.
151. Zeidler, R. Tests of efficiency in the rural and village schools of Santa Clara, Cal. *ElScJ*, 16: June, 1916, 542-555.

Drawing

152. Ayer, F. C. *The Psychology of Drawing.* Baltimore, Md., 1916, pp. 186.
153. Childs, H. G. Measurement of the drawing ability of 2,177 children in the Indiana city school systems by a supplemented Thorndike scale. *JEdPs*, 6: Sept., 1915, 391-408.
154. Lobsien, M. *Einige experimentelle Untersuchungen zu einigen Grundfragen der Kunsterziehung.* Langensalza, 1905.
155. Levinstein, S. *Untersuchungen ueber das Zeichnen der Kinder bis zum 14 Lebensjahr.* Leipzig, 1904.
156. Rugg, H. O. A scale for measuring freehand lettering. *JEdPs*, 6: Jan., 1915, 25-42.
157. Rugg, H. O. *A Scale for Measuring Freehand Lettering.* Chicago, Ill.
158. Thorndike, E. L. *Drawing Scale for Grades 5 to 8.* New York City.
159. Thorndike, E. L. The measurement of achievement in drawing. *TchCollRec*, 14: Nov., 1913, 345-383.

English Composition, Language, and Vocabulary

160. Ayres, L. P. *Spelling Vocabularies of Personal and Business Letters.* New York City, 1913, 14 pp.
161. Ballou, F. W. Scales for the measurement of English composition. *Harvard-Newton Bulletins*, No. 2, Cambridge, Mass., 1914, pp. 93.
162. Bliss, D. C. Some results of standard tests. *PsCl*, 6: March, 1912, 1-12.
163. Bonser F. G., Burch, L. H. and Turner, M. R. Vocabulary tests as measures of school efficiency. *ScSoc*, 2: Nov. 13, 1915, 713-718.
164. Boston, Department of Educational Investigation and Measurement, English: determining a standard in accurate copying. *Bulletin No. 6, School Document No. 2*, 1916, pp. 25.
165. Breed, F. S. and Frostic, F. W. Measuring English composition in the sixth grade. *ElScJ*, 17: Jan., 1917, 307-325.
166. Brownell, B. A test of the Ballou scale of English composition. *ScSoc*, 4: Dec. 16, 1916, 938-942.
167. Brownell, B. A criticism of recent attempts to measure language ability. *ScR*, 25: Sept., 1917, 495-503.
168. Earhart, G. and Small, J. English in the elementary school. *ElScJ*, 16: Sept., 1915, 32-48.
169. Faust, W. G. An effort to standardize descriptive theme-writing for the senior year of the high school. *EngJ*, 5: April, 1916, 257-271.
170. Green, J. A. Minimum essentials in elementary-school subjects: English. *JEPd*, 3: Dec., 1916, 380-383.
171. Hillegas, M. B. *Scale for English Composition by Young People.* New York City.

172. Hillegas, M. B. A scale for the measurement of quality in English composition by young people. *TchCollRec*, 13: Sept., 1912, 331-384.
173. Hillegas, M. B. *A Scale for the Measurement of English Composition by Young People.* New York City, pp. 60.
174. Hosic, J. F. The essentials of composition and grammar. *ScSoc*, 1: April 24, 1915, 581-587; also in *Fourteenth YbNaSoc, Pt. I*, 1915, 90-115.
175. Hudelson, E. Standards and measurements in English composition. *Second IndianaConfMes*, 13: 1915, 115-122.
176. Hudelson, E. Some achievements in the establishment of a standard for the measurement of English in the Bloomington, Indiana, schools. *EngJ*, 5: Nov., 1916, 590-597.
177. Jenkins, F. A test of the ability of children to use language forms. *JEdPs*, 6: June, 1915, 335-344.
178. Johnson, F. W. The Hillegas-Thorndike scale for the measurement of quality in English composition. *ScR*, 21: Jan., 1913, 39-49.
179. Johnson, R. I. The persistency of error in English composition. *ScR*, 25: Oct., 1917, 555-580.
180. Jones, R. G. Standard vocabulary. *Fourteenth YbNaSoc, Pt. I*, 1915, 37-43.
181. Kayfetz, I. A critical study of the Hillegas composition scale. *PdSe*, 21: Dec., 1914, 559-577.
182. Kayfetz, I. A critical study of the Harvard-Newton composition scales. *PdSe*, 23: Sept., 1916, 325-347.
183. Kelley, T. L. Individual testing with completion-test exercises. *TchCollRec*, 18: Sept., 1917, 371-382.
184. Kirkpatrick, E. A. A vocabulary test. *PopSciM*, 70: Feb., 1907, 157-164.
185. Noyes, E. C. and Miller, Edith. Progress in standardizing the measurement of composition. *EngJ*, 1: Nov., 1912, 532-536.
186. Reynolds, G. F. For minimum standards in English. *EngJ*, 4: June, 1915, 349-356.
187. Rice, J. M. Educational research: the results of a test in language; and English, the need of a new basis in education. *The Forum*, 35: Oct., 1903, 269-293 and March, 1904, 440-457. Reprinted in *Scientific Management in Education*, New York, 1913.
188. Sackett, L. W. Comparable measures of composition. *ScSoc*, 5: Feb. 24, 1917, 233-239.
189. Sears, Isabel and Diebel, Amelia. A study of the common mistakes in pupils' oral English. ElScJ, 17: Sept., 1916, 44-54.
190. Starch, D. and Elliott, E. C. Reliability of grading high-school work in English. *ScR*, 20: Sept., 1912, 442-457.
191. Stark, W. E. Measurement of eighth-grade composition. *ScSoc*, 2: Aug. 7, 1915, 208-216.
192. Stoddard, W. E. A comparison of the Hillegas and Harvard-Newton scales in English composition. *PdSe*, 23: Dec., 1916, 498-501.
193. Thompson, T. E. *Thompson's Minimum Essentials in Language.* Monrovia, Calif.
194. Thorndike, E. L. A scale for measuring the merit of English writing. *Sci*, 33: June 16, 1911, 935-938.
195. Thorndike, E. L. A scale for merit in English writing by young people. *JEdPs*, 2: 1911, 361-368.
196 Thorndike, E. L. *Preliminary Extension to the Hillegas Scale.* New York City.

197. Thorndike, E. L. *English Composition: 150 Specimens Arranged for Use in Psychological and Educational Experiments.* New York City.
198. Thorndike, E. L. Notes on the significance and use of the Hillegas scale for measuring the quality of English composition. *EngJ*, 2: Nov., 1913, 551-561.
199. Trabue, M. R. Some results of a graded series of completion tests. *ScSoc*, 1: April 10, 1915, 537-540.
200. Trabue, M. R. *Language Scales.* New York City.
201. Trabue, M. R. Completion tests for public-school use. *Fifteenth YbNaSoc*, Pt. I, 1916, 52-59.
202. Trabue, M. R. *Completion-Test Language Scales.* New York City, pp. 119.
203. Trabue, M. R. Supplementing the Hillegas scale. *TchCollRec*, 18: Jan., 1917, 51-84.
204. Trabue, M. R. *Nassau County Supplement to the Hillegas Scale.* New York City.
205. Ward, C. H. The scale illusion. *EngJ*, 6: April, 1917, 221-230.
206. Willing, M. H. Scale for measuring written composition. *KansasBurMes.*
207. Wolfe, C. S. The Topeka scale for judging compositions. *Illinois Association of Teachers of English Bulletin*, 9: No. 6: March 1, 1917, 1-12.

English Grammar

208. Betz, Annette and Marshall, Esther. Grammar based on errors. *EngJ*, 5: Sept., 1916, 491-500.
209. Charters, W. W. Minimal essentials in elementary language and grammar. *Sixteenth YbNaSoc, Pt. I*, 1917, 85-110.
210. Charters, W. W. and Miller, Edith. A course of study in grammar: based upon the grammatical errors of school children of Kansas City, Missouri. *University of Missouri Bulletin*, 16: No. 2: *Education Series* 9, Jan., 1915, pp. 45.
211. Maxwell, W. H. Report of the committee on tests and standards of efficiency in schools and school systems: English grammar. *ProcNaEdAs:* 1915, 565-573.
212. Rapeer, L. W. The problem of formal grammar in elementary education. *JEdPs*, 4: March, 1913, 125-137.
213. Starch, D. *Starch's English Grammar Tests.* Madison, Wis.
214. Starch, D. The measurement of achievement in English grammar. *JEdPs*, 6: Dec., 1915, 615-626.

French

215. Starch, D. *French Vocabulary and Reading Tests.* Madison, Wis.

Geography

216. Boston, Department of Educational Investigation and Measurement. Geography: a report on a preliminary attempt to measure some educational results. *Bulletin No. 5, School Document No. 14*, 1915, pp. 38.
217. Bagley, W. C. The determination of minimum essentials in elementary geography and history. *Fourteenth YbNaSoc, Pt. I*, 1915, 131-146.
218. Hahn, H. H. and Lackey, E. E. *Hahn-Lackey Geography Scale.* Wayne, Neb.
219. Starch, D. *Starch's Geography Test.* Madison, Wis.

220. Thompson, T. E. *Thompson's Minimum Essentials in Geography.* Monrovia, Calif.
221. Witham, E. C. A minimum standard for measuring geography. *AmScBdJ,* 1: Jan., 1915, 13-14.

Geometry

222. Metzler, W. H. Problems in the experimental pedagogy of geometry. *JEdPs,* 3: Dec., 1912, 545-560.
223. Stockard, L. V. and Bell, J. C. A preliminary study of the measurements of abilities in geometry. *JEdPs,* 7: Dec., 1916, 567-580.

German

224. Starch, D. *German Vocabulary and Reading Tests.* Madison, Wis.

Handwriting

225. Ashbaugh, E. J. Handwriting of the Iowa state children. *Extension Division Bulletin No. 15, University of Iowa,* March 1, 1916, pp. 24.
226. Ayres, L. P. *A Scale for Measuring the Quality of Handwriting of School Children. "Three Slant Edition."* New York City, 1912.
227. Ayres, L. P. *A Scale for Measuring the Quality of Handwriting of School Children.* Russell Sage Foundation, Bulletin E 113, 1912, pp. 16.
228. Ayres, L. P. *A Scale for Measuring the Quality of Handwriting of Adults.* New York City.
229. Ayres, L. P. *A Scale for Measuring the Quality of Handwriting of Adults.* Russell Sage Foundation E 138, pp. 11.
230. Ayres, L. P. *A Measuring Scale for Handwriting: "Gettysburg Edition."* New York City.
231. Boston, Department of Educational Investigation and Measurement. Penmanship. Determining the achievement of elementary school graduates in handwriting. *Bulletin No. 9, School Document No. 6,* 1916, pp. 43.
232. Breed, F. S. and Culp, V. An application and critique of the Ayres' handwriting scale. *ScSoc,* 2: Oct. 30, 1915, 639-647.
233. Breed, F. S. and Culp, V. Note on the relation of legibility and form in handwriting. *ScSoc,* 4: Dec. 2, 1916, 870-872.
234. Breed, F. S. and Down, E. F. Measuring and standardizing handwriting in a school system. *ElScJ,* 17: March, 1917, 470-484.
235. Courtis, S. A. *Courtis Standard Research Tests. Handwriting. Series W.* Detroit, Mich.
236. Freeman, F. N. *Freeman's Handwriting Scale.* Chicago, Ill.
237. Freeman, F. N. Some practical studies of handwriting. *ElScJ,* 14: Dec., 1913, 167-179.
238. Freeman, F. N. *The Teaching of Handwriting.* Boston, 1914, pp. 155.
239. Freeman, F. W. Handwriting. *Fourteenth YbNaSoc, Pt. I,* 1915, 61-77.
240. Freeman, F. N. An analytical scale for judging handwriting. *ElScJ,* 15: April, 1915, 432-441.
241. Freeman, F. N. Handwriting tests for use in school surveys. *ElScJ,* 16: Feb., 1916, 299-301.
242. Freeman, F. N. Handwriting. *Sixteenth YbNaSoc, Pt. I,* 1917, 60-72.
243. Graves, S. M. A study in handwriting. *JEdPs,* 7: Oct., 1916, 483-494.
244. Gray, T. C. The training of judgment in the use of the Ayres scale for handwriting. *JEdPs,* 6: Feb., 1915, 85-98.

245. Gray, T. C. A score card for the measurement of handwriting. ***Bulletin of the University of Texas,*** No. 37, 1915, pp. 50.
246. Harris, J. H. and Anderson, H. W. A study of handwriting in the public schools of Dubuque, Iowa. ***Bureau of School Measurements,*** Dubuque, 1916, pp. 10.
247. Johnson, G. L. and Stone, C. R. Measuring the quality of handwriting. *ElScJ,* 16: Feb., 1916, 302-315.
248. Johnson, J. H. A comparison of the Ayres and Thorndike handwriting scales (containing a table of equivalent values in the two scales). *North Carolina High-School Bulletin,* 7: Oct., 1916, 170-173.
248a. Kansas City, Mo. *Writing Scale.* Bureau of Research and Efficiency, Dec., 1917.
249. King, I. and Johnson, H. The writing abilities of the elementary and grammar-school pupils of a city system measured by the Ayres scale. *JEdPs,* 3: Nov., 1912, 514-520.
250. King, I. and Newcomb, R. E. Improvement in handwriting. ***EdAdm,*** 2: Oct., 1916, 493-502.
251. Lewis, E. E. The present standard of handwriting in Iowa training high schools. *EdAdm,* 1: Dec., 1915, 663-671.
252. Manuel, H. T. The use of an objective scale in grading handwriting (results of a study of the Ayres scale). *ElScJ,* 15: Jan., 1915, 269-278.
253. Manuel, H. T. Studies in handwriting. *ScSoc,* 5: March 17, 1917, 327-330.
254. Montgomery, L. J. Writing test. *Educator Journal,* 14: May, 1914, 479-483.
255. Nutt, H. W. Rhythm in handwriting. *ElScJ,* 17: Feb., 1917, 432-445.
256. Pintner, R. A comparison of the Ayres and Thorndike handwriting scales. *JEdPs,* 5: Nov., 1914, 525-536.
257. Rusk, R. R. A class experiment in scoring handwriting. *PEdPs,* 5: Sept., 1914, 417-418.
258. Sackett, L. W. Comparable measures of handwriting. *ScSoc,* 4: Oct., 21, 1916, 640-645.
259. Starch, D. The measurement of handwriting. *JEdPs,* 4: Oct., 1913, 445-464.
260. Starch, D. The measurement of efficiency in writing. *JEdPs,* 6: Feb., 1915, 106-114.
261. Thompson, Mary E. Psychology and pedagogy of writing. Baltimore, 1911, pp. 128.
262. Thorndike, E. L. Means of measuring school achievements in handwriting. *EdAdm,* 1: May, 1915, 300-305.
263. Thorndike, E. L. Handwriting. *TchCollRec,* 2: March, 1910, 1-93.
264. Thorndike, E. L. *Handwriting Scale for Grades 5 to 8.* New York City.
265. Thorndike, E. L. *Handwriting.* New York City, pp. 41.
266. Thorndike, E. L. Teachers' estimates of the quality of specimens of handwriting. *TchCollRec,* 15: Nov., 1914, 279-291.
267. Wilson, G. M. The handwriting of school children. *ElScTch,* 11: June, 1911, 540-543.
268. Witham, E. C. All the elements of handwriting measured. ***EdAdm,*** 1: May, 1915, 313-324.
269. Witham, E. C. A method of measuring handwriting. ***AmScBdJ,*** 48: May, 1914, 18-19 and 72-73.

History

270. Bell, J. C. and McCollum, D. F. A study of the attainments of pupils in United States History. *JEdPs*, 8: May, 1917, 257-274.
271. Foster, H. D. Adequate tests in history. *History Teacher's Magazine*, 5: April, 1914, 116-123.
272. Myers, G. C. Delayed recall in history. *JEdPs*, 8: May, 1917, 275-283.
273. Sackett, L. W. A scale in ancient history. *JEdPs*, 8: May, 1917, 284-293.
274. Starch, D. *Starch's American History Test. Series A.* Madison, Wis.
275. Starch, D. and Elliott, E. C. Reliability of grading work in history. *ScE*, 21: Dec., 1913, 676-681.

Latin

276. Brown, H. A. *Brown's Connected-Latin Test, Latin-Sentence Test, Latin-Grammar Test*, and *Latin-Vocabulary Test.* Oshkosh, Wis.
277. Brown, H. A. A study of ability in Latin. *Bulletin No. 2. Bureau of Research, New Hampshire Dept. of the Public Instruction*, Concord, N. H. (Published at State Normal School, Oshkosh, Wis.)
278. Hanus, P. H. Measuring progress in learning Latin. *ScE*, 24: May, 1916, 342-351.
279. Henmon, V. A. C. *Vocabulary and Translation Tests for Latin.* Madison, Wis.
279a. Henmon, V. A. C. The measurement of ability in Latin. Part I, Vocabulary. *JEdPs*, 8: Nov., 1917, 515-538.
280. Starch, D. *Latin Vocabulary and Reading Tests.* Madison, Wis.

Physics

281. Starch, D. *Test for Physics.* Madison, Wis.

Reading

282. Ballard, P. B. Norms of performance in reading. *JEPd*, 3: Dec., 1915, 153-161.
283. Bell, J. C. and Sweet, Itasca B. The reading interests of high-school pupils. *JEdPs*, 7: Jan., 1916, 39-45.
284. Boston, Department of Educational Investigation and Measurement. Standards in silent reading, with suggestions on how teachers may test their pupils in silent reading. *Bulletin No. 18, School Document No. 18*, 1916, pp. 24.
285. Bowden, Josephine. Learning to read. *ElScTch*, 12: Sept., 1911, 21-33.
286. Brown, H. A. The measurement of the efficiency of instruction in reading. *ElScTch*, 14: June, 1914, 477-490.
287. Brown, H. A. The measurement of ability to read. *New Hampshire Department of Public Instruction Bureau of Research, Bulletin No. 1*, pp. 57.
288. Brown, H. A. The significance of the measurement of ability to read. *Ed*, 36: May, 1916, 589-610.
289. Brown, H. A. *Brown's Silent Reading Test.* Oshkosh, Wis.
290. Childs, H. G. A half-year's progress in the achievement of one school system as measured by the Thorndike visual vocabulary test. *Fifteenth YbNaSoc, Pt. I*, 1916, 79-83.
291. Courtis, S. A. Standard tests in English. *ElScTch*, 14: April, 1914, 374-392.

292. Courtis, S. A. Standards in rates of reading. *Fourteenth YbNaSoc, Pt. I,* 1915, 44-58.
293. Courtis, S. A. The problem of measuring ability in silent reading. *AmScBdJ,* 54: May, 1917, 17-18, 81.
294. Courtis, S. A. *Courtis Standard Tests, Silent Reading, Test No. 2.* Detroit, Mich.
295. Gray, C. T. Types of reading ability as exhibited through tests and laboratory experiments. *Supplementary Educational Monographs,* 1, No. 5: Aug. 29, 1917, pp. 196. Chicago, Ill.
296. Gray, W. S. Selected bibliography upon practical tests of reading ability. *Fourteenth YbNaSoc, Pt. I,* 1915, 59-60.
297. Gray, W. S. Methods of testing reading. *ElScJ,* 16: Jan., 1916, 231-246 and Feb., 1916, 281-298.
298. Gray, W. S. *Gray's Reading Tests: Silent and Oral.* Chicago, Ill.
299. Gray, W. S. A study of the emphasis on various phases of reading instruction in two cities. *ElScJ,* 17: Nov., 1916, 178-186.
299a. Gray, W. S. A cooperative study in reading in eleven cities in northern Illinois. *ElScJ,* 17: Dec., 1916, 250-265.
300. Gray, W. S. The relation of silent reading to economy in education. *Sixteenth YbNaSoc, Pt. I,* 1917, 17-32.
300a. Gray, W. S. Studies of elementary school reading through standardized tests. *Supplementary Educational Monographs,* Chicago, 1917. No. 1, pp. 157.
301. Haggerty, M. E. Scales for reading vocabulary of primary children. *ElScJ,* 17: Oct., 1916, 106-115.
302. Haggerty, M. E. The ability to read: its measurement and some factors conditioning it. *IndianaUnSd,* 4: Jan., 1917, pp. 63.
303. Haggerty, M. E. *Visual Vocabulary Scale and Reading Tests for the Understanding of Sentences.* University of Minnesota, Minneapolis, Minn.
304. Haggerty, M. E. and Thomas, J. M. The reading attainments of college freshmen. *ScSoc,* 6: Aug. 25, 1917, 230-238.
305. Harris, J. H. and Anderson, H. W. Measuring primary reading in the Dubuque schools. Dubuque, 1916, pp. 23.
306. Hendricks, E. L. *A preliminary Report of a Study in Reading.* New York, 1911. pp. 30.
307. Hill, D. S. Experimental studies of achievement in reading. *Annual Report of the New Orleans Public Schools,* 1915-16, Pt. 3, 61-72.
308. Jones, R. G. *Jones' Scale for Teaching and Testing Elementary Reading.* Rockford, Ill.
309. Judd, C. H. Reading tests. *ElScTch,* 14: April, 1914, 365-373.
310. Judd, C. H. Report of the committee on tests and standards of efficiency in schools and school systems: reading tests. *ProcNaEdAs:* 1915, 561-565.
311. Judd, C. H. Reading. *Fifteenth YbNaSoc, Pt. I,* 1916, 111-119.
312. Kansas State Normal School, Bureau of Educational Measurements and Standards. *Standardized Tests in Silent Reading.* Emporia, Kan.
313. Kelley, T. L. Thorndike's reading scale Alpha 2 adapted to individual testing. *TchCollRec,* 18: May, 1917, 253-260.
314. Kelly, F. J. The Kansas silent reading test. *KansasBurMes,* 3: *New Series,* 4: 1915, pp. 37.
315. Kelly, F. J. The Kansas silent reading tests. *JEdPs,* 7: Feb., 1916, 63-80.
316. King, I. A comparison of slow and rapid readers. *ScSoc,* 4: Nov. 25, 1916, 830-834.

317. King, I. Comparison of the efficiency of slow and rapid readers. *ScSoc,* 6: Aug. 18, 1917, 203-204.
318. Llewelyn, E. J. Reading in the Mount Vernon, Indiana, city schools. *ElScJ,* 17: Oct., 1916, 123-127.
319. Lodge, R. C. and Jackson, J. L. Reproduction of prose passages. *PsCl,* 10: Oct., 1916, 128-145.
320. McGinnis, W. C. Reading. *Journal of Education,* 86: Sept. 6, 1917, 207-208.
321. Mead, C. D. Silent versus oral reading with one hundred sixth-grade children. *JEdPs,* 6: June, 1915, 345-348.
321a. Monroe, W. S. A report on the use of the Kansas Silent Reading tests with over one hundred thousand children. *JEdPs,* 8: 1917.
322. Moore, E. C. An experiment in teaching college students how to study. *ScSoc,* 2: July 17, 1915, 100-107.
323. Oberholzer, E. E. Testing the efficiency of reading in the grades. *ElScJ,* 15: Feb., 1915, 313-322.
324. Oberholzer, E. E. The effects of efficiency tests in reading on a city school system. *Fifteenth YbNaSoc, Pt. I,* 1916, 138-140.
325. Otis, A. S. Considerations concerning the making of a scale for the measurement of reading ability. *PdSe,* 23: Dec., 1916, 528-549.
326. Peters, C. C. The influence of speed drills upon the rate and effectiveness of silent reading. *JEdPs,* 8: June, 1917, 350-366.
327. Pintner, R. Oral and silent reading of fourth-grade pupils. *JEdPs,* 4: June, 1913, 333-337.
328. Pintner, R. and Gilliland, A. R. Oral and silent reading. *JEdPs,* 7: April, 1916, 201-212.
329. Richards, Alva M. and Davidson, P. E. Correlations of single measures of some representative reading tests. *ScSoc,* 4: Sept. 2, 1916, 375-377.
330. Schmidt, W. A. An experimental study in the psychology of reading. *Supplementary Educational Monographs,* 1: No. 2, April, 1917, pp. 126. Chicago.
331. Schmitt, Clara. School subjects as materials for tests of mental ability *ElScJ,* 15: Nov., 1914, 150-161.
332. Starch, D. *Starch's Silent Reading Tests.* Madison, Wis.
333. Starch, D. The measurement of efficiency in reading. *JEdPs,* 6: Jan., 1915, 1-24.
334. Starch, D. *Starch's English Reading Vocabulary Test.* Madison, Wis.
335. Thorndike, E. L. The measurement of ability in reading. *TchCollRec,* 15: Sept., 1914, 207-277.
336. Thorndike, E. L. Tests of oral and silent reading. *IndianaUnBu* 12: 1914, 37-51.
337. Thorndike, E. L. *Thorndike's Visual Vocabulary Reading Test.* New York City.
338. Thorndike, E. L. *Thorndike's Scale Alpha for Measuring the Understanding of Sentences.* New York City.
339. Thorndike, E. L. *Improved Scale for Measuring the Understanding of Sentences: Scale Alpha 2, Parts 1 and 2.* New York City.
340. Thorndike, E. L. An improved scale for measuring ability in reading. *TchCollRec,* 16: Nov., 1915, 31-53; and 17: Jan., 1916, 40-67.
341. Thorndike, E. L. *Improved Scale for Word Knowledge or Visual Vocabulary: Scale A 2 and Scale B.* New York City.
342. Thorndike, E. L. Measurement of achievement in reading: word knowledge. *TchCollRec,* 17: Nov., 1916, 430-454.
343. Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. *JEdPs,* 8: June, 1917, 323-332.

344. Thorndike, E. L. The understanding of sentences: a study of errors in reading. *ElScJ*, 18: Oct., 1917, 98-114.
345. Uhl, W. L. The use of the results of reading tests as a basis for planning remedial work. *ElScJ*, 17: Dec., 1916, 266-275.
346. Waldo, K. D. Tests in reading in Sycamore schools. *ElScJ*, 15: Jan., 1915, 251-268.
347. Wassen, A. W. Report of an experiment in the use of the Kansas silent reading test with Korean students. *EdAdm*, 3: Feb., 1917, 98-101.
348. Whipple, G. M. and Curtis, Josephine N. Preliminary investigation of skimming in reading. *JEdPs*, 8: June, 1917, 333-349.
349. Whitney, F. L. Measuring the value of first-grade readers. *AmScBdJ*, 53: Sept., 1916, 24 and 77-78.
350. Zeidler, R. Tests in silent reading in the rural schools of Santa Clara County, California. *ElScJ*, 17: Sept., 1916, 55-62.

Spanish

350a. Lewis, E. E. Tentative vocabulary tests for first-year Spanish. *Midland Schools*, 32: Sept., 1917, 6-8.

Spelling

351. Ayres, L. P. *A Measuring Scale for Ability in Spelling.* New York City.
352. Ayres, L. P. *A Measuring Scale for Ability in Spelling.* Bulletin E139, New York City, pp. 56.
353. Ballou, F. W. Measuring Boston's spelling ability by the Ayres spelling scale. *ScSoc*, 5: March 3, 1917, 267-270.
354. Bauer, N. *The Writing Vocabulary of Pupils of the New Orleans Public Schools.* Department of Superintendence, New Orleans, 1915.
355. Boston, Department of Educational Investigation and Measurement. Provisional minimum and supplementary lists of spelling words for pupils in Grades 1 to 8. *Bulletin No. 1, School Document No. 8:* 1914, pp. 48.
356. Boston, Department of Educational Investigation and Measurement. Spelling. Determining the degree of difficulty of spelling words. *Bulletin No. 4, School Document No. 10:* 1915, pp. 48.
357. Boston, Department of Educational Investigation and Measurement. Spelling: the teaching of spelling. *Bulletin No. 11, School Document No.17:* 1916, pp. 21.
358. Buckingham, B. R. Spelling ability: its measurement and distribution. *TchCollConEd*, 59: 1913, pp. 116.
359. Burnham, W. H. Hygiene and psychology of spelling. *PdSe*, 13: Dec., 1906, 474-501.
360. Cook, W. A. and O'Shea, M. V. *The Child and his Spelling.* Indianapolis, 1914, pp. 282.
361. Cornman, O. P. *Spelling in the Elementary School: an Experimental and Statistical Investigation.* Boston, 1902, pp. 98.
362. Courtis, S. A. *Teaching Spelling by Plays and Games.* Detroit, 1917, pp. 48.
363. Foster, F. M. The results of a recent spelling test at the University of Iowa. *ScSoc*, 5: April 28, 1917, 506-508.
364. Foster, W. T. The spelling of college students. *JEdPs*, 2: 1911, 211-215.
365. Gregory, B. C. The rationals of spelling. *ScHomeEd*, 27: Nov., 1907, 84-92, Dec., 1907, 123-129; also *ElScTch*, 8: Sept., 1907, 40-55.

366. Grupe, Mary A. A review of the pedagogical studies in the teaching of spelling. *Ed*, 34: Sept., 1913, 1-19.
367. Hill, D. S. Experimental studies in achievement in spelling. *Annual Report of the New Orleans Public Schools*, 1915-16, Pt. 3, 39-59.
368. Houser, J. D. The relation of spelling ability to general intelligence and to meaning vocabulary. *ElScJ*, 16: Dec., 1915, 190-199.
369. Houser, J. D. An investigation of the writing vocabularies of representatives of an economic class. *ElScJ*, 17: June, 1917, 708-718.
370. Jones, W. F. *Concrete Investigation of the Material of English Spelling.* Vermillion, S. Dak., 1914, pp. 27.
370a. Kansas City, Mo. *Preliminary Report of the Committee on Spelling.* Research Bulletin No. 2, 1916. pp. 47.
371. Lewis, E. E. Testing the spelling abilities of Iowa school children by the Buckingham spelling tests. *ElScJ*, 16: June, 1916, 556-564.
372. McFarland, W. H. Relation between spelling judgment and spelling ability. *Midland Schools*, 31: Sept., 1916, 19-21.
373. Monroe, W. *A Timed-Sentence Spelling Test. KansasBurMes:* Emporia Kan.
374. Otis, A. S. The reliability of spelling scales, involving a "deviation formula" for correlation. *ScSoc*, 4: Oct. 28, Nov. 4, 11, and 18, 1916, 677-683, 716-722, 750-756, and 793-796.
375. Pearson, H. C. The scientific study of the teaching of spelling. *JEdPs*, 2: 1911, 241-252.
376. Pryor, H. C. Spelling. *Fourteenth YbNaSoc, Pt. I*, 1915, 78-89.
377. Pryor, H. C. A suggested minimal spelling list. *Sixteenth YbNaSoc, Pt. I*, 1917, 73-84.
378. Rice, J. M. The futility of the spelling grind. *The Forum*, 23: Mar.-Aug., 1897, 163-172, 409-419. Reprinted in *Scientific Management in Education*, New York, 1913.
379. Sackett, L. W. Measuring a school system by the Buckingham spelling scale. *ScSoc*, 2: Dec. 11 and 18, 1915, 860-864 and 894-898.
380. Sackett, L. W. Disparity of spelling scales. *JEdPs*, 6: Dec., 1915, 627-630.
381. Scofield, F. A. Difficulty of Ayres' spelling scale as shown by the spelling of 560 high-school students. *ScSoc*, 4: Aug. 26, 1916, 339-340.
382. Scofield, F. A. An experiment in spelling in the Eugene High School. *ScSoc*, 5: March 10, 1917, 299-300.
383. Sears, J. B. Spelling efficiency in the Oakland schools. *ScSoc*, 2: Oct. 9 and 16, 1915, 531-537 and 569-574. Full report in *Board of Education Bulletin No. 9:* 1915, Oakland, Cal., pp. 79.
384. Starch, D.The measurement of efficiency in spelling and the overlapping of grades in combined measurements of reading, writing, and spelling. *JEdPs*, 6: March, 1915, 167-186.
385. Studebaker, J. W. *Results of an Investigation of Pupils' Ability to Spell.* Boston, 1916, pp. 37.
386. Suzzalo, H. The teaching of spelling. *TchCollRec*, 12: Nov., 1911, 2591330.
387. Suzzalo, H. *The Teaching of Spelling.* Boston, 1913, pp. 129.
388. Suzzalo, H. and Pearson, H. C. Comparative experimental teaching of spelling. *TchCollRec*, 13: Jan., 1912, 1-66.
389. Thorndike, E. L. Means of measuring school achievements in spelling. *EdAm*, 1: May, 1915, 306-312.
390. Tidyman, W. F. A critical study of Rice's investigation of spelling efficiency. *Pd8e*, 22: Sept., 1915, 391-400.

391. Tidyman, W. F. A descriptive and critical study of Buckingham's investigation of spelling efficiency. *EdAdm*, 2: May, 1916, 290-304.
392. Wallin, J. E. W. *Spelling Efficiency in Relation to Age, Grade, and Sex, and the Question of Transfer.* Baltimore, 1911, pp. 91.

DIVISION C

GENERAL REPORTS ON THE USE OF TESTS AND SCALES IN SCHOOLS

393. Ballou, F. W. Report of department of educational investigation and measurement. *Annual Report Superintendent of Schools Boston, School Document No. 17*, 1915, 84-118.
394. Ballou, F. W. Work of the department of educational investigation and measurement, Boston. *Fifteenth YbNaSoc, Pt. I*, 1916, 61-68.
395. Bobbitt, F. B. The work of the schools. Elementary schools. *Report of the School Survey of Denver, Col., Pt. 2*, 1916, 59-74.
395a. Bronner, Augusta F. *The Psychology of Special Abilities and Disabilities.* Boston, 1917. pp. 269.
396. Brown, G. A. Helps from the use of standard scales of attainment in school subjects. *ScHomeEd*, 33: April, 1914, 300-302.
396a. Chapman, J. C. and Feder, R. B. The effect of external incentives and improvement. *JEdPs*, 8: Oct., 1917, 469.
396b. Comin, R. Teachers' estimates of the ability of pupils. *ScSoc*, 3: Jan. 8, 1916, 67.
397. Courtis, S. A. *Courtis Standard Research Tests. Sixth Annual Accounting.* Also *Standard Scores in Silent Reading.* Detroit, 1917.
398. Courtis, S. A. Better teaching in arithmetic. *Proc. Harvard Teachers' Association*, 1913.
398a. Courtis, S. A. Measurement of the relation between physical and mental growth. *Amer. Phys. Educ. Rev.* 22: Nov., 1917.
399. Cubberley, E. P. Use of standard tests at Salt Lake City, Utah. *Fifteenth YbNaSoc, Pt. I*, 1916, 107-110.
400. Elliott, C. H. Variation in the achievements or pupils. A study of the achievements of pupils in the fifth and seventh grades, and in classes of different sizes. *TchCollConEd*, 72: 1914, pp. 114.
401. Haggerty, M. E. Elementary teachers and co-operative research. *IndianaUnBu*, 12: 1914, 94-110.
402. Harlan, C. L. A comparison of the writing, spelling, and arithmetic abilities of country and city children. *EdAdm*. 2: Nov., 1916, 560-572.
403. Judd, C. H. *Measuring the Work of the Public Schools.* The Survey Committee of the Cleveland Foundation, Cleveland, Ohio, 1916, pp. 290.
404. Kansas City, Mo., Bureau of Research and Efficiency, Public Schools. *Bulletin No. 1.*, Feb., 1916.
405. Kansas State Normal School. Bureau of Educational Measurements and Standards. Second and third annual reports, 1915-16 and 1916-17. *Study No. 6, New Series, 6, No. 7.* Emporia, Kansas.
406. Melcher, G. Studies by the bureau of research and efficiency of Kansas City, Mo. *Fifteenth YbNaSoc, Pt. I.*, 1916, 120-137.
407. Monroe, W. S. A teacher's handbook of educational measurements. *KansasBurMes.* 5: Emporia, Kan.
408. Monroe, W. S. *Educational Tests and Measurements.* New York City, 1917. pp. 309.
409. Morrison, J. C. The supervisor's use of standard tests of efficiency. *ElScJ*, 17: Jan., 1917, 335-354.
410. Scars, J. B. The efficiency of the instruction measured. *Report of a Survey of the School System of Salt Lake City, Utah*, 1915, 127-186.
411. Southington, Conn. *School Report*, 1917.

DIVISION D

LISTS OF TESTS AND SCALES

412. Alexander, C. List of investigations by members of the society of college teachers of education. *SclMon*, 5: 1914, 73-82.
413. Chapman, J. C. and Rush, Grace P. *The Scientific Measurements of Classsroom Products.* New York City, 1917, pp. 191.
414. Courtis, S. A. *A Manual of Instruction for Giving and Scoring the Courtis Standard Tests in the Three R's.* Detroit, 1914. pp. 128.
415. Cross, A. Weighing the scales. *EngJ*, 6: March, 1917, 183-191.
416. Gray, W. S. Descriptive list of standard tests. *ElScJ*, 17: Sept., 1916, 24-34.
417. Gray, W. S. Descriptive list of standard tests, 1917. *ElScJ*, 18: Sept. 1917, 56-70.
418. Holmes, H. W. and others. A descriptive bibliography of measurement in elementary subjects. *Harvard Bulletins in Education*, 5: June, 1917, pp. 43.
419. Johnston, J. H. A brief tabular history of the movement toward standardization by means of scales and tests of educational achievements in the elementary school subjects. *EdAdm*, 2: Oct., 1916, 483-492.
420. Knox, H. A. Measuring human intelligence. A progressive series of standardized tests used by the public health service to protect our racial stock. *Scientific American*, Jan. 9, 1915, 52-55.
421. Kohs, S. C. The Binet-Simon measuring scale for intelligence: an annotated bibliography. *JEdPs*, 5: April, May, and June, 1914, 215-224, 279-290, and 335-346.
422. Kohs, S. C. An annotated bibliography of recent literature on the Binet-Simon scale (1913-1917) Pt. I. *JEdPs*, 8: Sept. to Dec., 1917, 425-438, 488-502, 559-565, 609-618.
423. Manahan, J. L. (Compil.) A bibliography of educational surveys and tests. *University of Virginia Record Extension Series, 2:* Nov., 1916, 49-92.
424. Pyle, W. H. *The Examination of School Children. A Manual of Directions and Norms.* New York City, 1913. pp. 70.
425. Springer, I. (Ed.) Teachers' yearbook of educational investigations. *Division of Reference and Research, Bulletin No. I,* Department of Education, New York City, 1914, pp. 12.
426. Springer, I. (Ed.) Teachers' yearbook of educational investigations. *Division of Reference and Research, Publication No. 6,* Department of Education, New York City, 1915, pp. 75.
427. Springer, I. (Ed.) Teachers' yearbook of educational investigations. *Division of Reference and Research, Publication No. 14,* Department of Education, New York City, 1916. pp. 53.
428. Starch, D. *The Measurement of Efficiency in Reading, Writing, Spelling, and English.* Madison, Wis. 1914.
429. Starch, D. *Educational Measurements.* New York City, 1916, pp. 202.
430. Strayer, G. D. Standards and tests for measuring the efficiency of schools or systems of schools. *U. S. Bureau of Education Bulletin,* No. 13, 1913, pp. 23.
430a. Whipple, G. M. *Manual of Mental and Physical Tests.* Baltimore, Md., 1915, pp. 702.

DIVISION E

CORRELATIONS BETWEEN ABILITIES

431. Buckingham, B. R. Correlation between ability to think and ability to remember, with special reference to United States history. *ScSoc,* 5: April 14, 1917, 443-449.

431a. Chapman, J. C. Individual differences in ability and improvement and their correlation. *TchCollConEd,* No. 63, 1914.

432. King, Irving. The relationship of abilities in certain mental tests to ability as estimated by teachers. *ScSoc,* 5: Feb. 17, 1917, 204-209.

433. McCall, W. A. Correlation of some psychological and educational measurements, with special attention to the measurement of mental ability. *TchCollConEd,* 79: 1916, pp. 87.

434. McCall, W. A. Correlation of some psychological and educational measurements; with special attention to the measurement of mental ability. *ScSoc,* 5: Jan. 6, 1917, 24-30.

435. Simpson, B. R. Correlation of mental abilities. *TchCollConEd,* 53: 1912, pp. 122.

436. Starch, D. Correlations among abilities in school studies. *JEdPs,* 4: Sept., 1913, 415-418.

437. Terman, L. M. A trial of mental and pedagogical tests in a civil service examination for policemen and firemen. *Journal of Applied Psychology,* 1: March, 1917, 17-29.

438. Wissler, C. The correlation of mental and physical tests. PsR *Psychological Review Monograph Supplement,* 3, No. 6: June, 1901, New York, pp. 62.

439. Woolley, Helen T. and Fischer, Charlotte R. Mental and physical measurement of working children; studies from the laboratory of the vocation bureau, Cincinnati, Ohio, 1914. *The Psychological Monographs,* 18, No. 1: Dec., 1914, pp. 247.

DIVISION F

TEACHER MEASUREMENT

440. Anderson, D. A. Methods of measuring teaching efficiency. *ScSoc,* 3: April 15, 1916, 556-562.

441. Anderson, W. N. The selection of teachers. *EdAdm,* 3: Feb., 1917, 83-90.

442. Boyce, A. C. Qualities of merit in secondary school teachers. *JEdPs,* 3: March, 1912, 144-157.

443. Boyce, A. C. Methods for measuring teachers' efficiency. *Fourteenth YbNaSoc, Pt. 2,* 1915, pp. 83.

444. Coffman, L. D. The rating of teachers in service. *ScRMon,* No. 5: 1914 13-24.

445. Elliott, E. C. *Outline of a Tentative Scheme for the Measurement of Teaching Efficiency. Wisconsin State Department of Education.* 1910. Leaflet.

446. Elliott, E. C. Provisional plan for the measure of merit of teachers. *Wisconsin State Department of Education,* 1914.

447. Elliott, E. C. Methods and standards for determining teaching efficiency and the rating of teaching efficiency in other cities. *City School Supervision.* Yonkers, N. Y., 1914, Chaps. 9 and 10, 116-117.

448. Elliott, E. C. How shall the efficiency of teachers be tested and recorded? *ProcNaEdAs*, 1915, 469-473. Also in *EdAdm*, 1: May, 1915, 291-299.
449. Fichandler, A. A study in self-appraisal. *ScSoc*, 4: Dec. 30, 1916, 1000-1002.
450. Heffelfinger, J. B. The question as to the factor in teaching. *Tch*, 1, No. 10. State Normal School, Emporia, Kans.
451. Johnston, J. H. Teacher rating in large cities. *ScR*, 24: Nov., 1916, 641-647.
452. Johnston, J. H. Scientific supervision of teaching. *ScSoc*, 5: Feb. 17, 1917, 181-188.
453. Jones, E. S. A suggestion for teacher measurement. *ScSoc*, 6: Sept. 15, 1917, 321-322.
454. Kilpatrick, W. H. The rating of prospective teachers. *ScRMon*, No. 5: 1914, pp. 12.
455. Pittenger, B. F. Problems of teacher measurement. *JEdPs*, 8: Feb., 1917, 103-110.
456. Shiels, A. The rating of teachers in the New York City public schools. *ScSoc*, 2: Nov. 20, 1915, 752-754.
457. Sprague, H. A. Score card for rating student-teachers in training and practice. *PdSe*, 23: March, 1917, 72-80.
458. Springer, I. (Ed.) Teaching efficiency. *Teachers' Yearbook of Educational Investigations, Publication* No. 14; Bureau of Reference and Research, Department of Education, New York City, 1916.
459. Taylor, J. S. Measurement of educational efficiency. *Educational Review*, 44: Nov., 1912, 348-367.
460. Turner, E. A. Standards employed in the determination of teaching efficiency. *Normal School Quarterly*, Series 14, No. 58: Jan., 1916, Normal, Ill. pp. 32.
461. Witham, E. C. School and teacher measurement. *JEdPs*, 5: May, 1914, 267-278.

DIVISION G

ARTICLES ABOUT SURVEYS AND LISTS OF SURVEYS

462. Allen, W. H. Contrasted methods of surveying higher education. *ScSoc*, 5: Feb. 10, 1917, 171-173.
463. Ayres, L. P. School surveys. *ScSoc*, 1: April 24, 1915, 577-581.
464. Ayres, L. P. A survey of school surveys. *Second IndianaConfMes*, 13: 1915, 172-181.
465. Bobbitt, J. F. The school survey: finding standards of current practice with which to measure one's own schools. *ElScJ*, 15: Sept., 1914, 41-54.
466. Buckingham, B. R. A survey of the Gary schools and the pre-vocational schools in New York City. *Third IndianaConfMes*, 2: 1917, 119-125.
467. Buchner, E. F. School surveys. *Report of the U. S. Commissioner of Education*, 1, chapt. 24; 1914.
468. Buchner, E. F. Educational surveys, 1916. *Report of the U. S. Commissioner of Education*, 1, chapt. 21, 1916.
469. Elliott, E. C. State school surveys. *AmScBdJ*, 48: March, 1914, 9-10, and 62.

470. Frost, N. A statistical study of the public schools of the southern Appalachian mountains. *United States Bureau of Education Bulletin* No. 11, 1915, pp. 71.
471. Koos, L. V. The fruits of school surveys. *ScSoc,* 5: Jan. 13, 1917, 35-41.
472. Mahoney, J. Some foreign educational surveys. *United States Bureau of Education, Bulletin* No. 37, 1915, pp. 39.
473. Pintner, R. A mental survey of the school population of a village. *ScSoc,* 5: May. 19, 1917, 597-600.
474. Righter, L. The educational survey preparatory to the organization of vocational education. *TchCollRec,* 14: January, 1913, 3-42.
475. Smith, H. L. and Judd, C. H. Plans for organizing school surveys with a summary of typical school surveys. *Thirteenth YbNaSoc, Pt. 2,* 1914, pp. 85.
476. Strayer, G. D., Kandel, I. L., and Alexander, C. Selected bibliography of recent school surveys. *ScRMon,* No. 5: 1914, 62-72.
477. Thomson, F. A. College and university surveys. *ScSoc,* 5: June 23, 1917, 721-728.
478. Van Sickle, J. H., Ayres, L. P., and others. The investigation of the efficiency of schools and school systems. *ProcNaEdAs:* 1915, 379-402.
479. Witham, E. C. School measurement. *JEdPs,* 5: Dec., 1914, 571-588.

DIVISION H

CITY SURVEYS

480. Ashland, O. Constructive survey of the public school system of Ashland, Oregon. By F. C. Ayer, C. R. Frazier, and D. C. Sowers. *University of Oregon Bulletin,* 12: 1915, pp. 52.
481. Atlanta, Ga. *Report of Survey of the Department of Education.* By S. G. Lindholm. Bureau of Municipal Research, New York, Dec., 1912, pp. 24.
482. Atlanta, Ga. *Survey of the Atlanta Public Schools.* By Miss C. S. Parrish. State School Department, Atlanta, May, 1914, pp. 33.
483. Baltimore, Md. Report of the Commission appointed to study the system of education in the public schools of Baltimore. *United States Bureau of Education Bulletin,* 4: 1911, pp. 112.
484. Blaine, Wash. A survey of the public schools. By H. G. Lull and others. *University Extension Journal,* 1: July, 1914, University of Washington, pp. 165.
485. Bloomington, Ind. A survey of a public school system. By H. L. Smith. *TchCollConEd,* 82: 1917, New York, pp. 304.
486. Boise, Idaho. *Report of C. N. Kendall, Commissioner of Education of New Jersey, on the Boise Public Schools.* Board of Education, Boise, December 18, 1910, pp. 31.
487. Boise, Idaho. *Expert Survey of Public School System, Boise, Idaho.* By E. C. Elliott, C. H. Judd, and G. D. Strayer. 1913, pp. 31.
488. Boise, Idaho. *Special Report of the Boise Public Schools.* Board of Education, Boise. June, 1915, pp. 96.
489. Boston, Mass. *Report on the Boston School System.* Finance Commission of the City of Boston. Boston, 1911, pp. 234.
490. Boston, Mass. Report of a study of certain phases of the public school system of Boston, Mass. *Finance Commission of the City of Boston, Document* 87: 1916, pp. 219.

491. Boston, Mass. *Report on the Boston School Department: with Especial Emphasis on the Need for a Reorganization of Its Central Administrative System.* Finance Commission of the City of Boston, 1916, pp. 66.
492. Bridgeport, Conn. *Report of the Examination of the School System of Bridgeport, Conn.* Conducted by James H. Van Sickle. Board of Education, Bridgeport, Conn., 1913, pp. 129.
493. Brookline, Mass. *Educational Survey of the Public Schools.* By J. H. Van Sickle and others. The School Committee, Brookline, Mass., 1917, pp. 436.
494. Brooklyn, N. Y. *Report upon Divisions 4 and 5, Elementary Schools.* By William McAndrew. Reprinted by the World Book Company, Yonkers, N. Y., under title *The Public Health and its School.* 1917. pp. 76.
495. Buffalo, N. Y. *Examination of the Public School System of the City of Buffalo.* University of the State of New York, Albany, N. Y., 1916, pp. 208.
496. Butte, Mont. *Report of a Survey of the School System of Butte, Mont.* Board of School Trustees, Butte, Mont., 1914, pp. 163.
497. Chicago, Ill. *Report of Educational Commission of Chicago.* University of Chicago Press, Chicago, Ill., 1900, pp. 248.
498. Chicago, Ill. *Survey of Chicago Public Schools.* By E. F. Young. Chicago, Ill., 1914, pp. 257.
499. Chicago, Ill. *Recommendations for Reorganization of the Public School System of Chicago.* City Council, Chicago, Ill., 1917, pp. 83.
500. Cleveland, Ohio. *Report of the Educational Commission Appointed by the Board of Education to Examine into the Government, Supervision and Course of Study of the Cleveland Public Schools.* Board of Education, Cleveland, Ohio, 1906, pp. 120.
501. Cleveland, Ohio. *The Three R's. Improvement in Them. Place in Present Course of Study, Cleveland Public Schools.* Board of Education, Cleveland, 1909, pp. 30.
502. Cleveland, Ohio. *The Cleveland Education Survey.* L. P. Ayres, director. Survey Committee of the Cleveland Foundation, Cleveland, 1915-1916. 25 vols.
503. Dansville, N. Y. *A Study of the Dansville High School.* By J. M. Foster. Board of Education, Dansville, N. Y., 1915, pp. 108.
504. Denver, Col. *Report of the School Survey of School District No. 1, in the City and County of Denver.* School Survey Committee, Denver, Col., 1916.
505. East Orange, N. J. *Report of the Examination of the School System of East Orange, New Jersey.* Board of Education, East Orange, N. J., 1912, pp. 64.
506. East Windsor, Conn. *Educational Enquiry.* State Board of Education, Hartford, Conn., 1916, pp. 34.
507. Evansville, Ind. Report of the Evansville, Indiana, survey for vocational education. C. H. Winslow, director. Indiana State Board of Education. *Educational Bulletin, No. 19,* Indianapolis, Jan. 1, 1917.
508. Framingham, Mass. Survey of the Framingham Public Schools. By G. D. Strayer and others. *Annual Report for 1916 of the Board of School Committee and of the Superintendent of Schools,* Framingham, Mass., 34-84.
509. Gary, Ind. The public school system of Gary, Indiana. By W. P. Burris. *U. S. Bureau of Education Bulletin,* No. 18: 1914, Washington, D. C., pp. 49.
510. Glastonbury, Conn. *Educational Enquiry.* State Board of Education, Hartford, Conn., 1916, pp. 37.

511. Grafton, W. Va. *Report of the Survey of the Grafton City Schools.* By J. N. Deahl and others. Department of Schools, Charleston, W. Va., 1913, pp. 29.

512. Grand Junction, Col. *A Survey of the City Schools of Grand Junction, Colorado, District No. 1, Mesa County.* Board of School Directors, Grand Junction, Col., May, 1916, pp. 64.

513. Grand Rapids, Mich. *A School Survey of Grand Rapids, Mich.* Board of Education, Grand Rapids, Mich. 1916, pp. 510.

514. Great Neck, Long Island. *The Great Neck School Survey.* By G. D. Strayer and M. G. Neale. Great Neck Association, Great Neck, L. I., 1917, pp. 30.

515. Greenwich, Conn. *Report of the Special Committee and Town School Committee of Greenwich, Conn., with Reference to the Needs of the Town as to Enlarged School Accommodations.* Greenwich, Dec., 1911, pp. 11.

516. Greenwich, Conn. *The Book of the Educational Exhibit of Greenwich, Conn.* Greenwich, 1912, pp. 24.

517. Hammond, Ind. *Some Facts Concerning the People, Industries and Schools of Hammond: and a Suggested Program for Elementary Industrial, Prevocational and Vocational Education.* By R. J. Leonard. University of Indiana, Bloomington, Ind.

518. Harrisburg, Pa. *Report on the High School Situation in Harrisburg, Pa.* By H. Snyder. Board of Education, Harrisburg, 1912, pp. 23.

518a. Harrisburg, Pa. *A Report upon the Business and Financial Administration of the Harrisburg School District.* Bureau of Municipal Research, New York City.

519. Indianapolis, Ind. Report of the Indianapolis, Indiana, survey for vocational education, vol. 1. *Indiana State Board of Education, Educational Bulletin No. 21:* Jan., 1917.

519a. Janesville, Wis. *An Educational Survey of Janesville, Wis.* State Department of Education, 1918.

520. Leavenworth, Kans. Report of a survey of the public schools of Leavenworth, Kansas. *Bureau of Educational Measurements and Standards. Kansas State Normal School New Series,* 4, No. 2: Emporia, Kans., 1915, pp. 202.

557 [illegible] Angeles, Cal. *Report of the Advisory Committee to the Board of Education of the City of Los Angeles on Certain Aspects of the Organization and Administration of the Public School System.* Board of Education, Los Angeles, 1916, pp. 177.

522. Minneapolis, Minn. *A Vocational Survey of Minneapolis.* The Teachers' Club, Minneapolis, 1913, pp. 90.

523. Minneapolis, Minn. *A Series of Monographs on the Minneapolis Public Schools.* Board of Education, Minneapolis, Minn., 1916 and 1917.

524. Montclair, N. J. *Report on the Programme of Studies in the Public Schools of Montclair, N. J.* By P. H. Hanus. Cambridge, Mass., 1911, pp. 28.

525. Newburgh, N. Y. Public Schools. By F. Zeiger and Z. L. Potter. *The Newburgh Survey:* Russell Sage Foundation, New York, 1913, 5-20.

526. New Orleans, La. *Survey of Industries and Mechanical Occupations in New Orleans.* By the Division of Educational Research of the Public Schools. Summary of the report, *ScSoc,* 2: Sept. 18, 421-427; and Sept. 25, 1915, 461-466. Reprinted.

527. New York City, N. Y. *Final Report of Committee on School Inquiry, Board of Estimate and Apportionment.* The Committee, New York City, 1911-1913. 3 vols.

528. New York City, N. Y. *Outside Cooperation with the Public Schools of Greater New York; Report Submitted to the Board of Estimate's Committee on School Inquiry.* Abridged edition, Bureau of Municipal Research, New York, March, 1912, pp. 48.

529. New York City, N. Y. *A Digest of the New York School Inquiry.* Bureau of Municipal Research, New York City, 1913, pp. 68.

530. New York City, N. Y. *School Efficiency. A Constructive Study Applied to New York City, being a Summary and Interpretation of the Report on the Educational Aspects of the School Inquiry.* By P. H. Hanus. Yonkers, N. Y., 1913, pp. 121.

531. New York City, N. Y. *New York School Inquiry. Reply of the Superintendents.* By J. S. Taylor. New York, 1914, pp. 116.

532. New York City, N. Y. Help-your-own-school suggestions. Extracts from a Field Study of P. S. 188B Manhattan made at the request of Principal E. Mandel. Bureau of Municipal Research, *Municipal Research*, No. 31: Feb., 1914, pp. 32.

533. North Stonington, Conn. *Educational Enquiry.* State Board of Education, Hartford, Conn., 1916, pp. 29.

534. Oakland, Cal. Summary of a survey of the school department of Oakland. *Report No. 19, Tax Association of Alameda County*, Oakland, Cal., June, 1915, pp. 20.

535. Oakland, Cal. Report of a survey of the organization, scope, and finances of the public school system of Oakland, Cal. By E. P. Cubberley. *Board of Education Bulletin* No. 8, Oakland, Cal., June, 1915, pp. 48.

536. Ogden, Utah. *Report of the Ogden Public School Survey Commission.* By W. S. Deffenbaugh and others. State Department of Education, Salt Lake City, Utah, 1914, pp. 42.

537. Portland, Oregon. *Report of the Survey of the Public School System of School District No. 1. Multnomah County, Oregon, City of Portland.* Portland, The Committee, 1913, and Yonkers, N. Y., 1915, pp. 441.

538. Port Townsend, Wash. A survey of the Port Townsend public schools. H. G. Lull, director. *University of Washington Bulletin, University Extension Series No. 17, General Series No. 95*, Seattle, Wash., Aug., 1915, pp. 112.

539. Reading, Pa. Report on a survey of the board of education, by the New York Bureau of Municipal Research. *Report on a Survey. Municipal Departments and the School Districts*, Chamber of Commerce, Reading, 1914, 339-367.

540. Richmond, Ind. Report of the Richmond, Indiana, survey for vocational education. R. J. Leonard, director. *Indiana State Board of Education, Educational Bulletin*, 18: Indianapolis, Ind., Dec. 1, 1916.

541. Richmond, Ind. Survey of the Kindergartens of Richmond, Indiana. By Alice Temple. *Supplementary Educational Monographs*, 5: Chicago, Ill., 1917, pp. 58.

542. Richmond, Va. Vocational education. Under the direction of L. P. Ayres. *Bureau of Labor Statistics Bulletin No. 162*, Washington, D. C. Aug., 1915.

543. Rockford, Ill. *A Review of the Rockford Public Schools.* Board of Education, Rockford, Ill., 1915-1916, pp. 126.

544. St. Paul, Minn. *Classroom Instruction in St. Paul, Minn. Help-your-school Surveys.* Bureau of Municipal Research, New York, 1913.

545. Salt Lake City, Utah. *Report of a Survey of the Public School System of Salt Lake City, Utah.* Board of Education, Salt Lake City, June, 1915, pp. 324.

546. San Antonio, Texas. *A Survey of the San Antonio Public School System.* By J. F. Bobbitt. San Antonio, May, 1915, pp. 257.
547. San Francisco, Cal. *Some Conditions in the Schools of San Francisco.* School Survey Class, San Francisco, May, 1914, pp. 96.
548. Seymour, Conn. *Educational Enquiry.* State Board of Education, Hartford, Conn., 1916, pp. 26.
549. South Bend, Ind. *Superintendent's Report of the School City of South Bend, Indiana, and School Survey by the Department of Education of the University of Chicago.* Board of School Trustees, South Bend, Ind., Sept., 1914, pp. 219.
550. Springfield, Ill. *The Public Schools of Springfield, Illinois.* Under the direction of L. P. Ayres. Russell Sage Foundation, New York, 1914, pp. 152.
551. Syracuse, N. Y. *Report of Investigations for the Associated Charities of Syracuse, N. Y.* Bureau of Municipal Research, New York, 1912, pp. 12.
552. Waterbury, Conn. Waterbury public schools. *Help-your-school Surveys,* Bureau of Municipal Research, New York, 1913.

DIVISION I

STATE, COUNTY AND OTHER SURVEYS

553. Alabama. Alabama's public school system. A comparative study. *Department of Education Bulletin,* No. 55, Montgomery, 1916, pp. 32.
554. Alabama (Covington, Macon and Morgan Counties). An educational survey of three counties in Alabama. *Department of Education Bulletin* No. 43, Montgomery, 1914, pp. 179.
555. Arizona (Maricopa County). *Report of the School Survey Made in Maricopa County by the County Teachers' Association and the Arizona State Taxpayers' Association.* Arizona State Taxpayers' Association, 1917, pp. 8.
556. California (San Mateo County). Reorganizing a county system of rural schools. By J. H. Williams. *U. S. Bureau of Education Bulletin* No. 16, 1916, pp. 50.
557. Colorado. The rural and village schools of Colorado. An eight-year survey of each school district, 1906-13, inclusive. By C. G. Sargent. *Colorado Agricultural College, Series 14,* No. 5, 1914, pp. 106.
558. Colorado. A general survey of public high-school education in Colorado. *University of Colorado Bulletin,* 14: Oct., 1914, pp. 92.
559. Colorado. Report of an inquiry into the administration and support of the Colorado school system. *U. S. Bureau of Education Bulletin* No. 5, 1917, pp. 93.
560. Connecticut. *Report of the Special Educational Commission.* State Board of Education, Hartford, 1909, pp. 14.
561. Delaware (New Castle County). Mental status of rural school children. By E. H. Mullan. *Reprint No. 377 from Public Health Reports,* Washington, D. C., 1916.
562. Georgia (Bulloch, Clayton and Taliaferro, Jackson, Monroe, Morgan, Raban and Randolph Counties). *Educational Surveys.* By M. L. Duggan. Department of Education, Atlanta, 1914-16.
563. Illinois. Reports of the Illinois educational commission to the General Assembly of the state of Illinois. *Biennial Reports of the Superintendent of Public Instruction,* 1908-10, 506-612 and 1910-12, 30-34.

564. Illinois. *A Statistical Survey of Illinois Colleges.* By B. W. Brown. Council of Church Boards of Education, Chicago, 1917, pp. 78.
565. Illinois. *Report on Educational Administration.* By J. M. Mathews. *Report of the Economy and Efficiency Committee,* 403-485. Springfield, 1915.
566. Illinois. *A Survey of the Field and Service of McKendree College, Lebanon, Ill.* By H. W. Hurt. 1917, pp. 31.
567. Illinois. *Illinois School Survey; a Cooperative Investigation of School Conditions and School Efficiency.* L. D. Coffman, Director. Illinois State Teachers' Association, 1917, pp. 377.
568. Indiana (Greene County). *Educational Survey.* By D. C. McIntyre. Bloomington, Indiana, 1916, pp. 110.
569. Indiana (Jefferson County). Report of the Jefferson County survey for vocational education. *Educational Bulletin* No. 20, *Indiana Survey Series,* No. 5, Jan., 1917, pp. 87.
570. Indiana (Porter County). Rural school sanitation, including physical and mental status of school children in Porter County, Ind. By T. Clark, G. L. Collins, and W. L. Treadway. *U. S. Public Health Bulletin,* No. 77, 1916, pp. 127. *Mental Status* reprinted.
571. Iowa. Report on the hygienic conditions in Iowa schools. By I. King. *University of Iowa Extension Bulletin* No. 11, May, 1915, pp. 36.
572. Iowa. State higher educational institutions of Iowa. *U. S. Bureau of Education Bulletin,* No. 19, 1916, pp. 223.
573. Kansas. Survey of accredited high-schools and professional directory. By H. W. Josselyn. *University of Kansas Bulletin,* 15: July, 1914, pp. 252.
574. Maryland. Report of the Maryland educational commission. *Department of Education Bulletin,* No. 2, Jan., 1910, pp. 10.
575. Maryland. *Report on the Educational Situation in Maryland by a Special Committee of the Board of State Aid and Charities.* Maryland School for Boys, Loch Raven, Md., 1914, pp. 33.
576. Maryland. *Preliminary Report of the Maryland Country Life Commission.* 1914, pp. 32.
577. Maryland. *Public Education in Maryland. A Report to the Maryland Educational Survey Commission.* By A. Flexner and F. P. Bachman. General Education Board, New York, 1916, pp. 176.
578. Maryland (Montgomery County). An educational survey of a suburban and rural county. By H. N. Morse, E. F. Eastman, and A. C. Monahan. *U. S. Bureau of Education Bulletin,* No. 32, 1913, pp. 68.
579. Massachusetts. Report of the commission on industrial and technical education. *Teachers College Educational Reprints* No. 1, New York, 1906, pp. 196.
580. Michigan (Upper Peninsula). Report of the bureau of research of the Upper Peninsula Education Association. *Mining Journal,* Marquette, Sept., 1913, pp. 48.
581. Minnesota. The rural school system of Minnesota. A study in School Efficiency. By H. W. Foght. *U. S. Bureau of Education Bulletin,* No. 20, 1915, pp. 56.
582. Missouri (Saline County). A study of the rural schools of Saline County, Mo. By J. D. Elliff and A. Jones. *University of Missouri Bulletin,* 16, No. 22, Education Series No. 11, Columbia, Mo., Aug., 1915, pp. 32.
583. New York (Oswego County). *A Survey of the First Supervisory District of Oswego County, State of New York,* 1915-1916. By Mildred Pratt. pp. 114.

584. New York (Westchester County). *A Study of Local School Conditions.* By A. J. Inglis. Westchester County Research Bureau, White Plains, N. Y., 1912, pp. 29.
585. North Dakota. *Report of the Temporary Educational Commission to the Governor and Legislature of North Dakota.* Grand Forks, N. D. The State, 1912, pp. 61.
586. North Dakota. State higher educational institutions of North Dakota. *U. S. Bureau of Education Bulletin,* No. 27, 1916, pp. 204.
587. Ohio. *Report of the Ohio State School Survey Commission to the Governor.* State School Survey Commission, Columbus O., 1914, pp. 352.
588. Oregon (Lane County). A rural survey. By F. C. Ayer and H. N. Morse. *University of Oregon Bulletin,* New Series, 13, No. 14, Aug., 1916. Education treated 71-109.
589. Oregon. Report of a survey of the University of Oregon made by the U. S. Bureau of Education, 1915. *University of Oregon Bulletin* 13, No. 4, Dec., 1915, pp. 28.
590. Pennsylvania. *Report on Rural Schools.* Pennsylvania State Educational Association, Harrisburg, Dec., 1914, pp. 103.
591. Pennsylvania. The Pennsylvania state normal schools and public school system. By E. O. Holland. *TchCollConEd,* 55: New York, 1912, pp. 94.
592. Porto Rico. *A Study of Certain Social, Educational, and Industrial Problems in Porto Rico.* By M. Bloomfield. Boston, 1912, pp. 28.
593. Texas. *The Organization and Administration of a State's Institutions of Higher Education.* By A. Lefevre. Austin, 1914, pp. 524.
594. Texas. A study of rural schools in Texas. By E. V. White and E. E. Davis. *University of Texas Bulletin* No. 364, Extension Series No. 62, Oct., 1914, pp. 167.
594a. Texas. *The State Institutions of Higher Education in Texas.* By the Educational Campaign Committee. Austin, Texas, 1912.
594b. Texas. *A Survey of Rural Schools in Travis County, Texas.* By E. E. Davis, Austin, Texas, 1916.
595. Vermont. *Secondary Education in Vermont.* By R. McFarland. Middlebury. 1912, pp. 42.
596. Vermont. Education in Vermont. *Carnegie Foundation for the Advancement of Teaching Bulletin* No. 7, Pts. 1 and 2, 1914, pp. 214.
596a. Vermont. *Report of the Commission to Investigate the Educational System and Conditions of Vermont.* Brattleboro, 1914.
597. Virginia. Report of the Virginia education commission to the General Assembly of the Commonwealth of Virginia. *Senate Document* No. 3, 1912, pp. 102.
598. Washington. A survey of educational institutions of the state of Washington. *U. S. Bureau of Education Bulletin,* No. 26, 1916, pp. 228.
599. Wisconsin. *Report of the Commission upon the Plans for the Extension of Industrial and Agricultural Training.* Madison, 1911, pp. 135.
600. Wisconsin. *Preliminary Report on Conditions and Needs of Rural Schools in Wisconsin.* State Board of Public Affairs, Madison, 1912, pp. 92.
601. Wisconsin. *Preliminary Report of the Committee of Fifteen Appointed to Investigate Educational Needs and Conditions in Wisconsin.* C. P. Cary, Madison, 1912, pp. 30.
602. Wisconsin. *Report upon the Survey of the University of Wisconsin.* State Board of Public Affairs, Madison, 1914, pp. 957.
602a. Wisconsin. *Conditions and Needs of Wisconsin Normal Schools.* State Board of Public Affairs, 1914.

603. Wisconsin. *Survey Summary re University of Wisconsin.* Wisconsin Efficiency Bureau, Madison, 1915, pp. 105.
604. Wisconsin. *The Milwaukee County School of Agriculture and Domestic Economy.* By Walter Matscheck. Madison, 1916, pp. 73.
605. Wyoming. Educational survey of Wyoming. By A. C. Monahan and Katherine M. Cook. *U. S. Bureau of Education Bulletin,* No. 29, 1916, pp. 120.
606. United States. *A Comparative Study of the Public School Systems in the Forty-Eight States.* Russell Sage Foundation, New York, 1912, pp. 32.

NATIONAL ASSOCIATION
OF
DIRECTORS OF EDUCATIONAL RESEARCH

CONSTITUTION

ARTICLE I

Name.—The name of this organization shall be the National Association of Directors of Educational Research.

ARTICLE II

Object.—The object of the Association shall be: (1) the formation of independent departments of educational research in all systems of public instruction, and (2) the promotion of the practical use of educational measurements in all educational research having for its object the improvement of the efficiency of the educational administration, supervision or teaching.

ARTICLE III

Membership.—SECTION 1. In general, membership in the Association shall be restricted to those who are actively and mainly engaged in research work having for its direct purpose the evaluation of the products of educational training or the improvement of the efficiency of educational teaching, supervision or administration.

SEC. 2. Membership in the Association may be either regular or associate. Regular members shall have any and all of the rights and privileges of the Association, including the right to vote, to hold office, and to appear upon any formal or informal program of the Association. Associate members are not eligible for office, have no vote and may not appear in any formal program of the Association, or otherwise represent the Association in public meetings, without special invitation of the Executive Committee; but they are to receive all bulletins of the Association, to be notified of all meetings or other activities and no distinction is to be made in any informal meeting or program between them and regular members.

SEC. 3. Any person holding the position of Director of, or supervising, a Department of Educational Research in an educational institution, or any immediate assistant of such director, shall be eligible for full membership.

SEC. 4. Any person actively engaged in research work in education, but holding some educational position other than in a department of research shall be eligible for associate membership.

SEC. 5. The Executive Committee, through the Secretary, shall, if necessary, ask all applicants for membership to state their positions, duties and past achievements in measurement work, and decision as to eligibility shall be made by the Executive Committee. In all cases where the applicant holds two or more positions, one of which has to do with educational research, the decision of the Executive Committee shall be made in accordance with the intent of the qualifications for membership as outlined above. Regular members who have not contributed to the bulletins of the Association during the year shall automatically become associate members at the annual meeting next following.

ARTICLE IV

Dues and Assessments.—There shall be no regular annual dues. The expenses incidental to carrying on the work of the Association shall be met by an assessment voted at each regular annual meeting.

ARTICLE V

Officers.—The officers shall consist of a president, a vice-president, and a secretary-treasurer. These officers shall be elected at the regular annual meeting of the Association. Their duties shall be those usually performed by such officers.

ARTICLE VI

Executive Committee.—There shall be one executive committee of five members consisting of the officers and the two preceding presidents whose duty shall be the conduct of the business of the Association between meetings.

ARTICLE VII

Meetings.—The time and place of holding the annual meeting shall be determined by a vote of the Association. Special meetings of the Association or of the executive committee may be called by the president, and must be called by him whenever requested by a majority of the executive committee.

ARTICLE VIII

Amendments.—Changes in this constitution may be made at any annual meeting of the Association by the affirmative vote of two-thirds of the members present.

MEMBERSHIP

Officers, 1917-18

President.....S. A. Courtis

Vice-president.....B. R. Buckingham

Secretary.....George W. Melcher, Library Bldg., Kansas City, Mo.

Executive Committee

F. W. Ballou (President 1915-16), W. S. Monroe (President 1916-17), S. A. Courtis, B. R. Buckingham, George W. Melcher.

Honorary Members

Hon. J. M. Rice, Germantown, Pa.
Dr. Paul H. Hanus, Harvard College, Cambridge, Mass.
Dr. E. L. Thorndike, Teachers College, Columbia University, New York City.
Dr. Charles H. Judd, University of Chicago, Chicago, Ill.

Regular Members

Allison, Samuel B., Asst. Superintendent in charge of Standards and Statistics, Chicago, Ill.

Anderson, Homer W., Educational Research Public Schools, 508 City Hall, Omaha, Neb.

Ashbaughm, Ernest J., Educational Service, Extension Division, University of Iowa, Iowa City, Iowa.

Averill, William A., State Department of Education, Albany, New York.

Ayres, Leonard P., Director Division of Education, Russell Sage Foundation, New York City.

Ballou, Frank W., Director Department of Educational Investigation and Measurements, Mason Street, Boston, Mass.

Bright, Ira J., Director of Research and Efficiency, Public Schools, Topeka, Kansas.

Brueckner, Leo J., Department of Educational Research and Martindale Normal School, Detroit, Mich.

Buckingham, B. R., Educational Statistics State Board of Education, Madison, Wis.

Burns, Allen T., Director Cleveland Foundation, Cleveland, Ohio.

Clark, Earle, Statistician, Russell Sage Foundation, New York City.

Courtis, S. A., Director Department of Educational Research 82 Eliot Street, Detroit, Michigan.

Davenport, Harry L., Director Division of Research and Statistics, Schnectady, N. Y.

DeVoss, J. C., Assistant, Bureau of Standards and Measurements State Normal School, Emporia, Kansas.

Dickson, Virgil E., Acting Director of Reference and Research, Public Schools, Oakland, Cal.

Flemming, Cecile White, Assistant Bureau of Educational Measurements, State Department of Education, Madison, Wis.

Fordyce, Charles, Dean Teachers College and Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Neb.

Haggerty, M. E., Director of Bureau of Cooperative Research, University of Minnesota, Minneapolis, Minn.

Halsey, G. D., Director of Educational Research and Vocational Guidance, Atlanta, Ga.

Hebden, Edwin, Director Bureau of Statistics and Research, 714 Euclid Avenue, Roland Park Branch P. O., Baltimore, Md.

Kelly, F. J., Dean School of Education and Director of Bureau of School Service, University of Kansas, Lawrence, Ks.

Lane, Robert E., Bureau of Research & Efficiency, Los Angeles, Cal.

Mackay, William A., Director of Research, Buffalo, N. Y.

Melcher, George, Director of Bureau of Research and Efficiency, Library Building, Kansas City, Mo.

Monroe, Walter S., Director of Bureau of Educational Standards and Measurements, State Normal School, Emporia, Kansas.

Nifenecker, Eugene A., Assistant Director Division of Reference and Research, 500 Park Avenue, New York City.

O'Hern, J. P., Assistant Superintendent of Schools and Director of Efficiency Bureau, Rochester, N. Y.

Phelan, W. W., Director of School of Education and Measurements and Efficiency, University of Oklahoma, Norman, Okla.

Race, Henrietta V., Director Psychological Clinic, Board of Education, Louisville, Ky.

Richardson, Arthur H., Russell Sage Foundation, New York City.

Richardson, J. Wylie, Educational Research Director, Hibbing, Minnesota.

Sharpleigh, F. E., Research Secretary, Buffalo Public Educational Association, Buffalo, N. Y.

Sutton, Clarence W., Director of Reference and Research, Cleveland, Ohio.

Talbert, Wilford E., Assistant in charge Bureau of Educational Investigations, 677 Melville Avenue, Palo Alto, Cal.

Theisen, W. W., Supervisor of Educational Measurements, State Department of Education, Madison, Wis.

Welles, J. B., Assistant Director of Reference and Research, Cleveland, Ohio.

Invitation

The Association urges all persons professionally engaged in educational research having for its purpose the improvement of the efficiency of school work, to join the Association and lend the influence of their names and positions to the efforts of the Association to stimulate the growth of the movement.

ANNOUNCEMENT OF YEARBOOKS AND EXPLANATION OF MEMBERSHIP IN THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

The purpose of the National Society is to promote the investigation and discussion of educational questions. Anyone who is interested in receiving its publications may become a member. The *Yearbooks* are issued in several Parts each year and are discussed at the annual meeting, which is held in February at the same time and place as the meeting of the Department of Superintendence of the National Education Association. There are two types of membership, associate and active. Associate members pay $1.00 annually and receive one copy of each *Yearbook*. Active members pay $2.00 annually, receive two copies of each *Yearbook*, and are eligible to vote and hold office in the Society.

The *Yearbooks* deal in a practical way with fundamental current issues in instruction and school administration. The *Seventeenth Yearbook* (calendar year 1918) will comprise Part I, to contain the "Third Report of the Committee of the National Education Association on the Economy of Time," and Part II, to contain thirteen chapters on "The Measurement of Educational Products," prepared by the National Association of Directors of Educational Research. Both Part I and Part II may be expected in February, 1918.

Orders for *Yearbooks for 1917 or earlier* or for *single parts of the Yearbook for 1918* are handled directly as commercial sales, by the Public School Publishing Co., Bloomington, Illinois, at the rates indicated on the cover of this monograph. To obtain the entire *Yearbook for 1918* as a member of the Society, pin your check or postal order to the following slip, properly filled out, and mail to the Secretary now.

APPLICATION FOR MEMBERSHIP

To GUY M. WHIPPLE
Secretary of the National Society for the Study of Education
The University of Illinois, Urbana, Illinois

Please enroll me as an { active / associate } member

I inclose { $2.00 as payment of active dues for calendar year 1918 / $1.00 as payment of associate dues for calendar year 1918 }

Name .....

Address to which to mail *Yearbooks* .....
.....